ENCYCLOPEDIA OF

Physical Science and Technology

THIRD EDITION

Molecular Biology

AP

Table of Contents
(Subject Area: Molecular Biology)

Article	*Authors*	Pages in the Encyclopedia
Cell Death (Apoptosis)	*Masato Enari*	Pages 541-554
Chromatin Structure and Modification	*Fyodor D. Urnov and Alan P. Wolffe*	Pages 809-829
DNA Testing in Forensic Science	*Moses S. Schanfield*	Pages 589-602
Gene Expression, Regulation of	*Göran Akusjärvi*	Pages 501-516
Immunology-Autoimmunity	*K. Michael Pollard and Eng M. Tan*	Pages 679-691
Ribozymes	*Alessandra Poggi and John J. Rossi*	Pages 253-261
Translation of RNA to Protein	*R. A. Cox and H. R. V. Arnstein*	Pages 31-51

Cell Death (Apoptosis)

Masato Enari
Imperial College School of Medicine at St. Mary's

I. Overview
II. Death Factors and Their Receptors
III. Apoptotic Proteases, Caspases
IV. Signal Transduction of Death Factor-Mediated Apoptosis
V. Cell-Free System in Apoptosis
VI. Apoptotic DNase, CAD and Its Inhibitor, ICAD
VII. Molecular Mechanism of CAD and ICAD
VIII. Physiological DNA Fragmentation and Phagocytosis of Apoptotic Cells
IX. Perspectives

GLOSSARY

Apoptosis The typical process in physiological cell death that is accompanied by nuclear and cytoplasmic condensation, fragmentation of cell bodies, chromosomal DNA fragmentation, loss of mitochondrial function, and alterations of cell membrane composition. It is distinct in these regards from necrosis. The term was created by Wyllie and Kerr.

Caspases Cysteine proteases, some of which are activated during apoptosis. Some caspases are involved in processing of cytokines.

Cell-free system A biochemical technique to be reconstituted cellular events as *in vitro* reaction.

Death factors Proteins belonging to the tumor necrosis factor (TNF) superfamily that occur on the cell surfaces as membrane-bound factors and in the extracellular compartment as soluble factors. They kill target cells expressing their corresponding receptors or to attenuate killing activity.

DNases Enzymes possessing DNA-cleaving activity. Some DNases participate in chromosomal DNA degradation during apoptosis.

DNA fragmentation Chromosomal DNA from apoptotic cells gives rise to a ladder pattern on agarose gels, due to multimeric nucleosomal units (~180 base pairs).

Death receptors Proteins belonging to TNF receptor family that occur on cell surface, and mediate killing by effector cells expressing their cognate ligands.

Intracellular signal transduction The cellular machinery that mediates external signals including hormones, neurotransmitters, cell growth, differentiation and death factors, or stress to their ultimate targets.

Phagocytosis The process by which phagocytes, such as macrophages and neutrophils, engulf useless and unnecessary cells.

APOPTOSIS, or programed cell death, plays an important role in the development of organisms and in the maintenance of homeostasis. The failure of apoptotic programs causes various diseases. So far, many genes regulating apoptosis have been identified and the molecular mechanism of apoptosis is being clarified. In this article, I will discuss cell death elicited through death receptors.

I. OVERVIEW

Homeostasis in multicellular organisms is based on a balance between life and death of cells. Apoptosis was recognized as a phenomenon distinct from necrosis by Wyllie and Kerr in 1972. In the necrotic process, swelling of cells precedes their explosion and results in the release of intracellular components that may be toxic to other cells. In apoptosis, the dying cells exhibit nuclear and cytoplasmic condensation, fragmentation of cell bodies, chromosomal DNA fragmentation into nucleosomal units, loss of mitochondrial function, and alterations of cell membrane composition (Fig. 1). Subsequently, apoptotic cells are engulfed by phagocytes and neighboring cells, and are recycled. Most cells suffering physiological cell death undergo the apoptotic process and the superfluous or harmful cells generated during the developmental process are removed by apoptosis. For example, apoptosis occurs in tail resorption, neuronal network formation, clonal deletion of

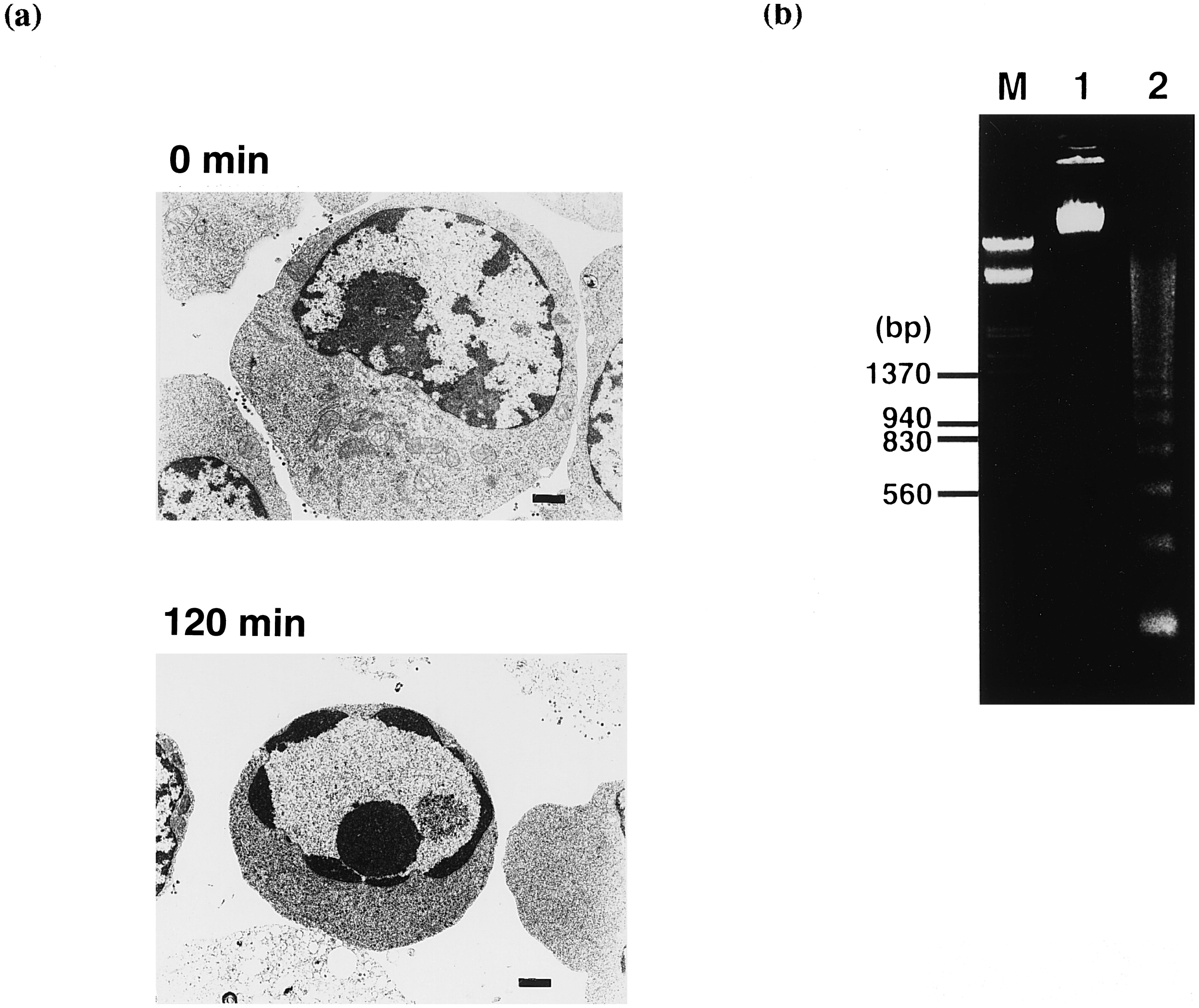

FIGURE 1 (a) Fas-induced apoptosis in lymphoid cells (WR19L cells overexpressing Fas). The cells were incubated with 0.5 μg/ml of an agonistic anti-Fas antibody at 37°C for 120 min and their ultrastructure was examined under a transmission electron microscope. The electron micrograph of untreated cells is shown in the upper panel. Bars, 1 μm. (b) Chromosomal DNA of growing cells (lane 1) or dying cells (lane 2) was run through a 1.5% agarose gel. M indicates molecular weight markers.

immature and autoreactive T cells, Wolffian and Mullerian duct regression during sexual development, tumor regression, and in elimination of virus-infected cells. Furthermore, it has been suggested that apoptosis occurs in many diseases such as cancer, fulminant hepatitis, acquired immune deficiency syndrome (AIDS), diabetes mellitus, and neurodegenerative disorders such as Alzheimer's disease and prion disease.

Growth and differentiation of cells are strictly regulated by factors such as cytokines and low-molecular-weight compounds such as steroid hormones. These factors are generally bound to the corresponding receptors in order to transduce the appropriate cell signals, to promote growth and differentiation. On the other hand, apoptotic cell death is aggressively controlled by a number of polypeptides, so-called death factors exposed at the cell surface or circulating in the body as soluble factors in some situations. Growth and differentiation factors act via transcriptional regulation through the activation of a series of protein kinases. On the other hand, death factors execute apoptosis through the activation of caspases, and many proteins essential for cell survival are degraded by these activated caspases. It is currently believed that apoptotic death is due to the degradation of many functional proteins by caspases. Cell-free systems for the study of apoptosis have been established and facilitate our understanding of the underlying molecular mechanisms. Several factors involved in apoptotic pathways have been identified and part of these pathways has been revealed by biochemical approaches based on cell-free apoptosis system. Regulatory mechanisms for apoptosis include: the CED-3/caspase family proteases and CED-4/Apaf-1 family which act as executors of apoptosis; CED-9/Bcl-2 family (including anti-apoptotic and pro-apoptotic factors) which act as regulators of apoptosis, and many factors which contribute to apoptotic morphological changes have been identified. Here, I shall focus on the currently proposed molecular mechanisms of death receptor activity and caspase activation. Moreover, I will also discuss the DNase responsible for apoptotic DNA fragmentation, both *in vitro* and *in vivo,* and finally the mechanism by which apoptotic cells are cleared.

II. DEATH FACTORS AND THEIR RECEPTORS

Many cells have death receptors on the surface of their plasma membranes and apoptosis is triggered by their cognate ligands. Death receptors belong to the superfamily of tumor necrosis factor (TNF) receptors. Most consist of a cysteine-rich extracellular domain, a membrane-spanning domain, and a cytoplasmic domain containing a death domain (DD) that is required for transducing apoptotic signals to cells. Six DD-containing death receptors, namely Fas/Apo1/CD95, type I TNF receptor (TNFRI), DR3/Apo3/WSL-1/TRAMP, DR4/TRAIL-R1 (TNF-related apoptosis-inducing ligand receptor-1), DR5/TRAIL-R2/TRICK2/KILLER and DR6 have been identified so far. In addition, there are DD-less death receptors such as type II TNF receptor (TNFRII), CD27, CD30, CD40, and lymphotoxin-β receptor. Among these, TNFRII, CD27, and CD30 induce the expression of both TNF and TNFRI by their cognate ligands, and subsequent induction of apoptosis appears to occur.

On the other hand, death factors are a member of the TNF family and exist on the cell surface or as soluble factors. Most death factors are synthesized as type II-membrane proteins and subjected to shedding by membrane-associated metalloproteinases in order to generate soluble forms of death factors. A well-established mechanism for shedding of death factors is seen in the case of TNF. Membrane-bound TNF (memTNF) is cleaved at the outer cellular membrane by a membrane-spanning protease, so-called TNF α-converting enzyme (TACE), which is a member of metalloproteinase-disintegrin (ADAM) family. memTNF is superior to soluble TNF (sTNF) in activating TNFRII in various cellular responses, including T-cell proliferation, inflammation, and cytotoxicity. These results imply that memTNF regulates cellular responses via restricted cell-to-cell interaction under physiological conditions and that sTNF may attenuate TNFRII-mediated responses. The shedding mechanism for other death factors may be similar to that for TNF. Like in shedding of TNF, the membrane form of Fas ligand (memFasL) is also cleaved by an unknown metalloproteinase other than TACE present on the plasma membranes. In addition, the soluble form of Fas ligand (sFasL) inhibits mFasL-mediated apoptosis in human peripheral blood T lymphocytes *in vitro.*

It is believed that death receptors are activated by ligand-induced trimerization, as opposed to the activation of growth factor receptors, which occurs by dimerization. Most death factors, but not all, are present as trimers. X-ray structural analyses have revealed that TNF-α, TNF-β, CD40L, and TRAIL are homotrimeric proteins. Among these, TRAIL has unique characteristics. TRAIL requires a zinc ion for biological activity and selectively induces apoptosis in mouse tumor cells but not in normal cells. Moreover, administration of soluble TRAIL to mice implanted with human tumors causes effective reduction of tumor size without any injury of normal tissues. These results suggest that TRAIL may be applicable as an anti-cancer drug. However, a recent paper reported that TRAIL induces apoptosis in human hepatocytes, indicating that substantial liver toxicity might result if

TRAIL were used in human cancer therapy. Before it can be used in a clinical setting, it would be necessary to understand the molecular mechanisms by which TRAIL transduces signals from its receptor in human hepatocytes.

III. APOPTOTIC PROTEASES, CASPASES

In the execution phase of apoptosis, caspase proteases are activated by various apoptotic stimuli. Caspases belong to the cysteine protease family and were originally identified as homologues of the *ced*-3 (cell death abnormal) gene product (an executor of apoptosis in *Caenorhabditis elegans*). Indeed, most caspases induce apoptosis if they are overexpressed in growing cells and cell death can be blocked by caspase-specific inhibitors. Therefore, caspases are accepted as executors for apoptosis in mammals and this pathway appears to be conserved between species.

Caspases are synthesized as zymogens and the active enzymes are generated by proteolytic cleavage of these caspase precursors. X-ray crystallographic studies have shown that active caspases are heterotetramers consisting of two large subunits (∼20 kDa) and two smaller (∼10 kDa) subunits. So far, 14 caspases have been cloned from mammalian sources and divided into three subgroups based on primary structure, phylogenetic analysis, and substrate specificity. Caspases cleave proteins on the carboxyl side of aspartic acid with sequence specificity. For example, caspase-1 shows preferential cleavage at Asp in the Tyr–Val–Ala–Asp sequence, whereas caspase-3 cleaves at latter Asp in the Asp–Glu–Val–Asp. Caspases are classified as initiator, effector and cytokine-releasing caspases, according to their functions. Initiator caspases include caspase-2, -8, -9, and -10. These have large prodomains, which interact with specific adapter molecules to convert the precursor to the active form. Effector caspases, including caspase-3, -6, -7, and -14, have short prodomains and are activated by active initiator caspases. This regulatory mechanism is known as a protease cascade. Other caspases found in mammals appear to serve as cytokine-releasing enzymes.

Gene disruption experiments have revealed the physiological functions of individual caspases *in vivo*. (1) Caspase-3- and -9-deficient mice show embryonic lethal phenotypes with neuronal hyperplasia, indicating that caspase-3 and -9 play an important role in neuronal development. Moreover, their embryonic fibroblasts are resistant to staurosporine-, etoposide-, UV-, and dexamethasone-induced apoptosis but not to Fas-mediated apoptosis. (2) Caspase-8-null mice are embryonic lethal and established cells from these mice are resistant to cell death through death receptors including Fas, DR3, and TNFRI, whereas these cells are sensitive to staurosporine-induced apoptosis. (3) Caspase-2 is required for the formation of female germ cells and loss of the caspase-2 gene renders B cells resistant to granzyme B-induced cell death. (4) Patients suffering from autoimmune lymphoproliferative syndrome (ALPS) type II, which display immune regulatory defects, have missense mutations in caspase-10. Caspase-10 mutants have less caspase activity and block Fas- and TRAILR1-mediated apoptotic pathways. (5) Both caspase-$1^{-/-}$ and caspase-$11^{-/-}$ mice are resistant to endotoxic shock, such as results from LPS treatment, and their gene products are required for the production of cytokines: caspase-1 is needed for secretion of IL-1α and caspase-11 for secretion of IL-1α and β. As discussed above, each caspase is important for both various developmental and pathological processes. Knockout analyses for other caspase genes are still under investigation and will reveal individual *in-vivo* functions for each caspase in the near future.

IV. SIGNAL TRANSDUCTION OF DEATH FACTOR-MEDIATED APOPTOSIS

As described earlier, Fas- or TNFRI-mediated apoptosis is triggered by binding of the cognate death ligands (Fig. 2). In general, apoptosis induced by most death factors can still occur in the presence of inhibitors of protein or RNA synthesis, suggesting that the death factor-mediated apoptotic processes proceed without *de novo* synthesis of either proteins or RNAs unlike the processes involved in growth and differentiation, and that all of components for apoptotic signaling are constitutively present in cells. There is an essential domain required for transduction of the death signals into cells, so-called death domain (DD), found in the cytoplasmic region of both Fas and TNFRI. Immunoprecipitation analyses have suggested that a protein complex, designated as the death-inducing signaling complex (DISC), is recruited to the cytoplasmic domain of Fas following the interaction of Fas with Fas ligand (Fig. 2). Using the yeast-two hybrid technique with the cytoplasmic region of Fas or TNFRI as baits, several molecules that specifically bind to the cytoplasmic region of these receptors have been discovered. FADD (Fas-associating protein with death domain)/MORT-1 is a small adapter protein with a molecular mass of 26 kDa and a death domain at the its C-terminus. FADD is recruited to trimerized cytoplasmic region of Fas and binds to Fas via interactions between the death domains (Fig. 2). Nuclear magnetic resonance (NMR) studies have shown that the death domain of Fas consists of six antiparallel, amphipathic α-helices arranged in a novel fold. Because there are many charged groups from amino acids on the protein

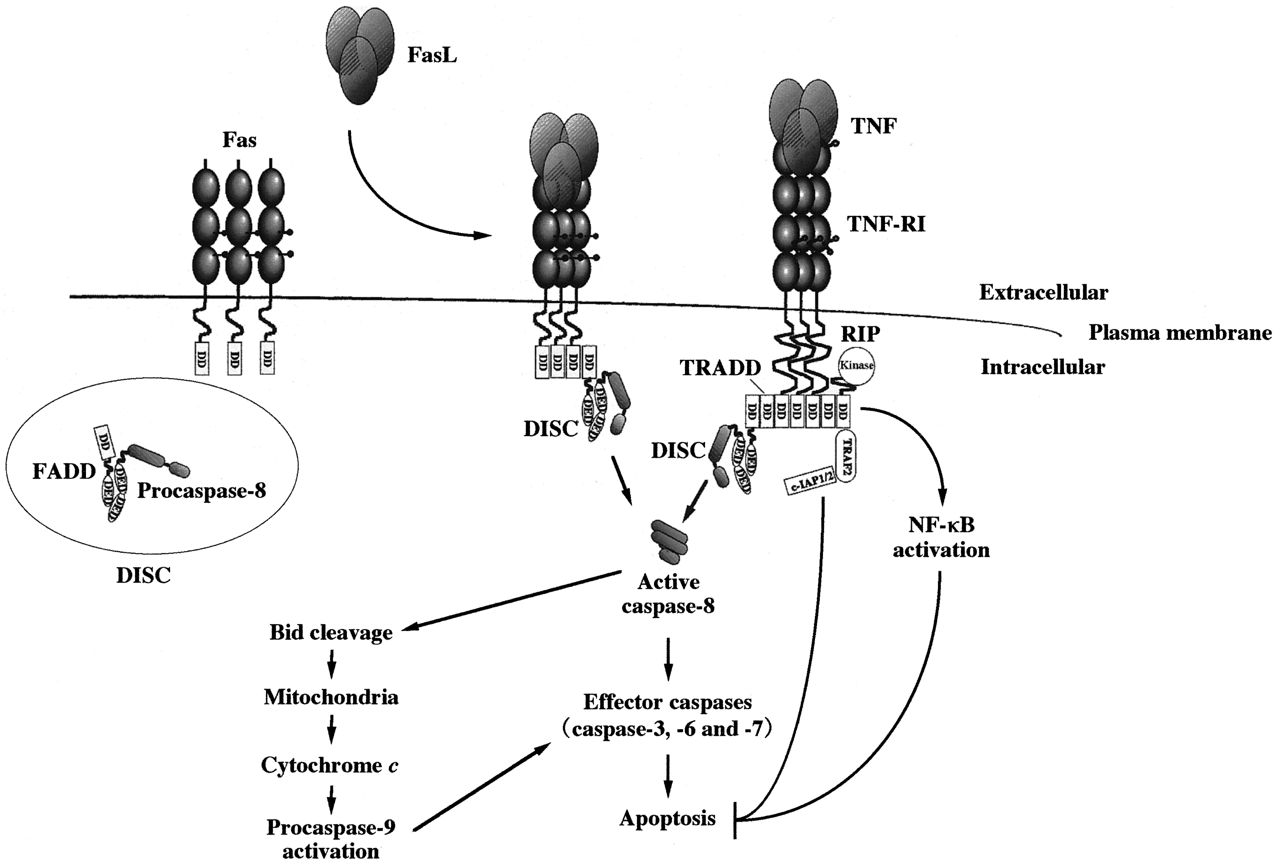

FIGURE 2 Signal transduction of Fas- and TNFRI-mediated apoptosis.

surface, the binding between death domains may be mediated by ionic interactions. Deletion mutant experiments have shown that the N-terminal region of FADD but not its death domain is necessary for transducing death signals. In addition, the FADD with deletion of the N terminus works as a dominant-negative mutant against the Fas-mediated system. Therefore, this region has been designated as death effector domain (DED). Similarly, the DD-possessing protein, TRADD (TNFRI-associated death domain protein), has been found as an adapter molecule that specifically binds to TNFRI. Unlike FADD, TRADD does not have DED, but is essential for mediating apoptosis induced by TNFRI. Subsequently, it has been shown that TRADD binds to FADD via DD interactions (Fig. 2). That is, TRADD is recruited to the cytoplasmic region of trimerized TNFRI that consequently binds FADD. Thus, Fas and TNFRI utilize the same transducer and share the apoptotic machinery downstream of FADD. The signal through TNFRI is more complex than that of Fas because of the diversity of signals from TNFRI. For example, RIP (receptor-interacting protein), originally identified as a Fas-binding protein with DD motif, preferentially binds to TRADD and is involved in the activation of transcriptional factor NF-κB that is responsible for stimulating the proliferation of thymocytes (Fig. 2). Although RIP has a kinase domain at the N terminus, this kinase domain is dispensable for activating NF-κB. Several analyses indicate that NF-κB appears to protect against TNFRI-mediated apoptosis via up-regulation of c-IAP2 (cellular inhibitor of apoptosis protein 2). Up-regulated c-IAP2 appears to bind TRAF2 (TNF receptor-associated factor 2), which can bind RIP and inhibit apoptotic signaling from TNFRI (Fig. 2). Both negative and positive apoptotic signals from TNFRI might be necessary for fine tuning the decision to die or not.

Two independent approaches involving the yeast two-hybrid experiments with DED of FADD as a bait and purification of factors binding to the cytoplasmic region of Fas, have revealed that procaspase-8 binds to the DED. Procaspase-8 contains two DED motifs at the N terminus through which it binds to FADD and a caspase homologous region at the C terminus (Fig. 2). Indeed, DISC contains both FADD and procaspase-8, and recruited procaspase-8 is processed to the active enzyme near the

inner plasma membrane. Several lines of evidences suggest that procaspase-8 may be proteolytically processed to the active form by simple oligomerization. However, the precise mechanism is still unclear and may require unknown additional factor(s) that may be in DISC or on the inner plasma membrane, may be required to generate active caspase-8 more efficiently. Once caspase-8 is activated in cells, it cleaves effector procaspases located downstream of it such as caspase-3, -6, and -7 in order to amplify the apoptotic signal (Fig. 2). Finally, cells show a variety of apoptotic features such as described above, due to cleavage of more than 100 substrate proteins which contributes to DNA fragmentation, loss of mitochondrial function, and the maintenance of nuclear structure and plasma membrane.

An alternative apoptotic pathway via mitochondria has been found by the *in vitro* reconstitution assay (described in section V). Based on a dATP/ATP-inducible cell-free system, three factors responsible for processing procaspase-3 have been purified and identified. These include Apaf-1, which is found as a mammalian homologue of CED-4 known to be another executor of apoptosis in *C. elegans*, cytochrome *c* (Apaf-2) that is mainly present in the mitochondrial intermembrane space, and procaspase-9 (Apaf-3). Cytochrome *c* is released from the intermembrane space of mitochondria into the cytoplasm during apoptosis induced by a variety of apoptotic stimuli, including DNA-damaging agents, protein kinase inhibitors (staurosporine), and death receptors (Fig. 3). Moreover, the release of cytochrome *c* is blocked by anti-apoptotic proteins belonging to the Bcl-2 family such as Bcl-2 and Bcl-xL (Fig. 3). Recent observations have indicated that cytochrome *c*-deficient mice show an embryonic lethal phenotype with defects in oxidative phosphorylation. In addition, their embryonic fibroblasts are resistant to stresses such as UV irradiation, serum withdrawal and staurosporine. The mechanism by which cytochrome *c* is released from mitochondria by apoptotic stimuli remains elusive, although some hypotheses, including opening of a specific channel for cytochrome *c*, alteration of the permeability transition pore (PTP) that regulates inner mitochondrial membrane potential, or swelling and subsequent rupture of the outer mitochondrial membrane, have been proposed.

Apaf-1 possesses a region homologous to the procaspase-prodomain, known as the caspase-recruiting domain (CARD) at the N terminus, a region homologous to CED-4 in the middle part and WD-40 repeat structure that appears to be involved in protein–protein interactions. The released cytochrome *c* interacts with two cytosolic proteins, Apaf-1 and procaspase-9, and dATP/ATP in the cytoplasm to form a complex known as the apoptosome (Fig. 3). Apaf-1 binds to procaspase-9 via CARD motifs and procaspase-9 is processed using both Apaf-1 and energy from dATP/ATP hydrolysis to convert procaspase-9 to the active form. Cytochrome *c* appears to play a role in overcoming the inhibition of the Apaf-1 active site masked by the WD-40 repeat. Once activated, effector caspases downstream of caspase-9 are sequentially activated (Fig. 3).

It has recently been shown that there are two cell types with different sensitivity to Fas signaling induced by an agonistic anti-Fas antibody. In type I cells, procaspase-8 is rapidly activated following receptor engagement, whereas in type II cells the activation of procaspase-8 is delayed, although both type I and type II cells show similar kinetics of Fas-mediated apoptosis and loss of mitochondrial function. In addition, Bcl-2 inhibits apoptosis in type II but not type I cells. Why do these cells die in a similar manner to each other regardless of the amount of active caspase-8 and yet show different blocking activity by Bcl-2? One explanation is that there are two distinct pathways in Fas-mediated apoptosis, a direct pathway from caspase-8 to effector caspase in type I cells and a caspase-8-mediated mitochondrial pathway in type II cells. It is likely that Bcl-2 is only able to block the latter pathway. However, this hypothesis is still controversial in view of the report that physiological Fas ligand but not an agonistic antibody kills both types of cells similarly regardless of the Bcl-2 expression level. These results might depend on the efficiency of precise trimerization of Fas. The evidence that there are two distinct pathways in the Fas system has come from the analyses of Bid-deficient mice. If Bid, a Bcl-2 family member, is cleaved by caspase-8, truncated Bid (tBid) can translocate from the cytosol to mitochondria, subsequently, cytochrome *c* is released, executing apoptosis. Administration of an agonistic anti-Fas antibody to wild-type mice *in vivo* causes death with hepatocellular apoptosis and haemorrhagic necrosis in the liver within 3 hours, whereas Bid-deficient mice survive treatment with this antibody. In addition, hepatocytes from Bid$^{-/-}$ mice are resistant to this antibody *in vitro,* suggesting that the tBid-mediated pathway via mitochondria predominantly works in hepatocellular apoptosis induced by agonistic Fas antibody. It will be interesting to determine whether or not Apaf-1$^{-/-}$ and caspase-9$^{-/-}$conditional knockout mice (because of the embryonic lethal phenotype) show a similar response as Bid-deficient mice to the administration of the anti-Fas antibody. To investigate the actual signaling pathways activated by the physiological Fas ligand in vivo, it will be necessary to administer soluble Fas ligand to Bid-deficient mice.

Some death receptors also activate caspase proteases by ligation with their cognate ligands. Like in TNFRI signaling, DD regions of activated death receptors including

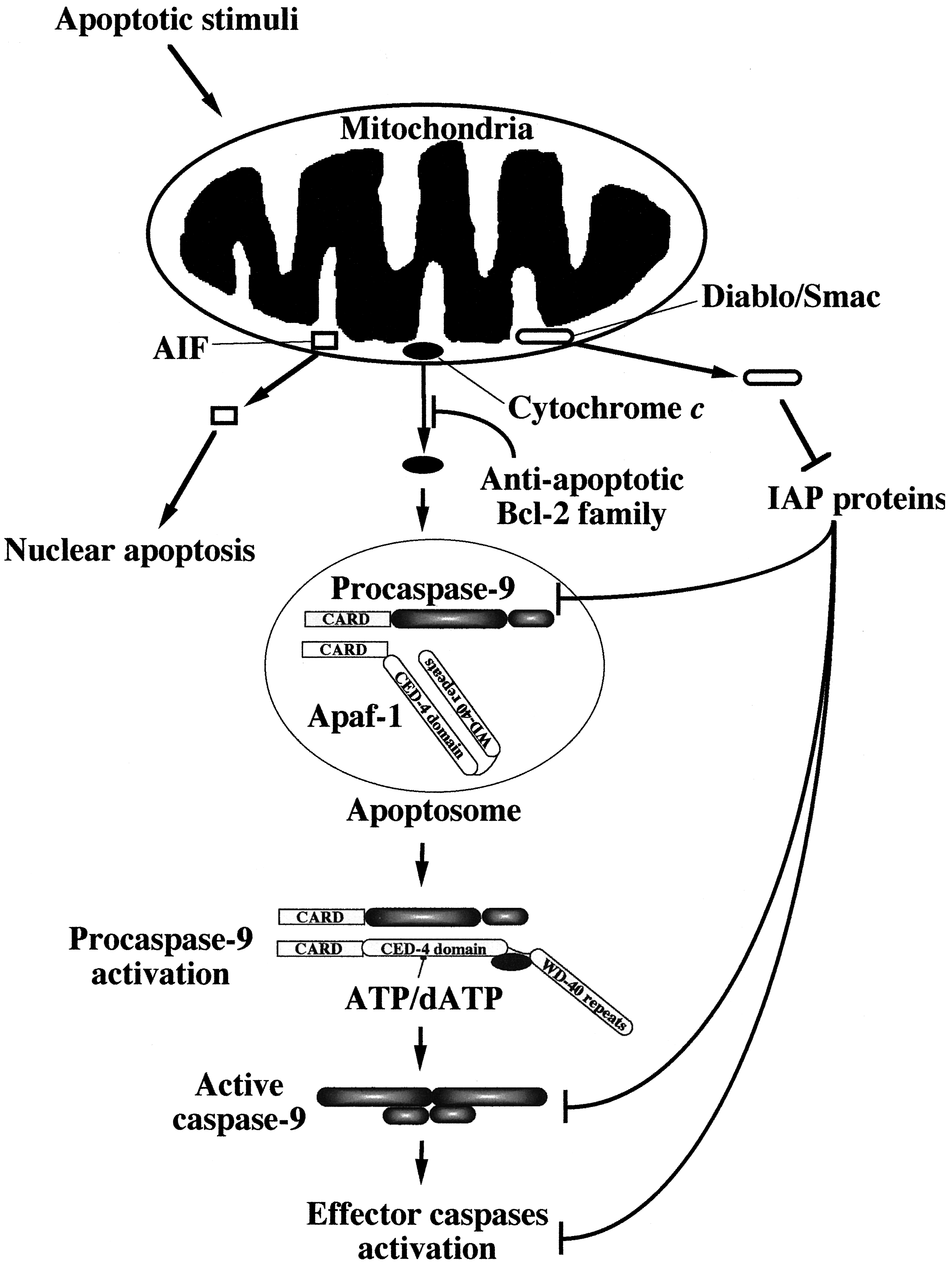

FIGURE 3 Apoptotic pathway via mitochondria.

DR3 and TRAILR2 recruit DISC (that consists of FADD and procaspase-8) via the DD of TRADD. Unlike Fas, TNFRI, and DR3, TRAILR1 can transmit apoptotic signals and activate caspase-8 in FADD-deficient fibroblast cells. This suggests that TRAILR1-mediated apoptosis is caspase-8-dependent and FADD-independent, although an unidentified FADD-like adapter molecules cannot be ruled out at present. The response mediated by DR6 has not, as yet, been well characterized.

In conclusion, ligand-induced trimerized death receptors activate an apical protease, procaspase-8, via FADD or an unidentified functional homologue. Activated caspase-8 can sequentially activate effector caspases responsible for cleaving a variety of death substrates such as ICAD/DFF45 (discussed below), lamin, fodrin, and poly(ADP-ribose) polymerase. In addition to the direct pathway from initiator caspases to effector caspases through Fas, active caspase-8 can transduce death signals into mitochondria through the translocation of caspase-8-cleaved Bid from cytosol to mitochondria. Uptake of tBid into mitochondria promotes the release of cytochrome *c* to the cytosol by unknown mechanisms. Once cytochrome *c* is released, it can generate active caspase-9 via formation of the apoptosome. Thus, death receptors appear to use two different pathways for particular tissues in some situations. The two different pathways ensure the death signals are amplified and target cells killed.

V. CELL-FREE SYSTEM IN APOPTOSIS

The establishment of cell-free systems has provided us with detailed insight into the cognate molecular mechanisms of various cellular functions, including general/specific transcriptional regulations, RNA editing, protein synthesis and protein degradation, and overall metabolic pathways. The well-established biochemical hallmark of apoptosis is chromosomal DNA degradation resulting in multimers of 180 bp-nucleosomal units. Neither protein- nor RNA-synthesis inhibitors block Fas-mediated apoptosis, suggesting that all of the components for apoptotic induction through Fas are present in cells in latent forms.

A cell-free system for apoptosis was first established by exposing isolated nuclei to various extracts and monitoring nucleosomal DNA fragmentation. That is, when isolated nuclei from healthy cells were treated with extracts from dying cells (but not from growing cells), they showed apoptotic features, including nuclear morphology with peripheral condensation of chromatin and nucleosomal DNA fragmentation. This cell-free system can be reproduced using the extracts from cells subjected to different apoptotic stimuli, including UV radiation and treatment with ceramide (N-acyl-erythro-sphingosine). In addition, extracts from cells activated by Fas in the presence of a caspase inhibitor do not cause DNA fragmentation in the cell-free reaction. Indeed, extracts from proliferating cells, in the presence of recombinant active caspase-3, induce nuclear apoptosis, indicating that the factors responsible for apoptotic DNA fragmentation are downstream of the caspase cascade. It emerged that these factors form a complex composed of heterodimeric proteins of caspase-activated DNase (CAD)/DNA fragmentation factor 40 (DFF40)/caspase-activated nuclease (CPAN) and an inhibitor of CAD (ICAD)/DFF45 (discussed in the following).

As described above, if dATP is mixed with extracts from proliferating cells, the extracts have the ability to induce nuclear apoptosis mediated by caspase-3 activation in a cell-free reaction. Using this system, Wang's group have purified the factors responsible for processing procaspase-3 and identified three gene products, namely Apaf-1, cytochrome *c* and procaspase-9. Thus, biochemical approaches using cell-free systems for apoptosis has demonstrated several important aspects in the field of apoptosis.

During apoptosis, biochemical procedures have frequently shown loss of mitochondrial function. In order to clarify the regulations of mitochondrial apoptosis, isolated mitochondria have been used. During the early stages of study in this field, the function of mitochondria in apoptotic processes was overlooked due to no apparent changes on morphology. However, the observations that Bcl-2 (an anti-apoptotic protein) is abundant in the mitochondria, and that apoptotic processes are often accompanied by a decrease in the mitochondrial membrane potential and by mitochondrial swelling led to a more detailed investigation of mitochondrial function during apoptosis. Furthermore, apoptogenic factors, such as cytochrome *c* and apoptosis-inducing factor (AIF) that had been discovered as a nuclear apoptosis-inducing flavoprotein using a cell-free system, are released from the intermembrane space of mitochondria into the cytosol to induce apoptosis. Recently, another factor secreted from mitochondria during apoptosis, called Diablo/Smac, has been reported. This factor suppresses the functions of proteins belonging to the inhibitor-of-apoptosis (IAP) family, inhibiting the protease activity of caspases including caspase-3, -7 and -9. Cell death is likely to be accelerated as a result of inhibition of IAP. Thus, there appear to be various steps to ensure cell killing. How is the loss of mitochondrial function induced by the various apoptotic stimuli? For example, Bax (Bcl-2-associated X protein), which belongs to a member of the pro-apoptotic Bcl-2 family, induces apoptosis. When added individually to purified mitochondria-free cytosol, neither mitochondria nor Bax can individually induce the activation of caspases that lead to nuclear apoptosis. In

contrast, the addition of both Bax and mitochondria triggers the release of cytochrome *c* from mitochondria to cytosol, which activates the caspase cascade. In addition, it has been shown that Bax directly triggers cytochrome *c* release from mitochondria in a cell-free reaction. It has been proposed that Bax may release cytochrome *c* from mitochondria by modulating voltage-dependent anion channel activity and/or causing limited permeabilization of the mitochondrial outer membrane although this mechanism is still controversial. The apoptotic signals from various stimuli mediated by mitochondria should be investigated, dissected, and reproduced by biochemical approaches using a cell-free system described here.

VI. APOPTOTIC DNase, CAD AND ITS INHIBITOR, ICAD

A biochemical hallmark of apoptosis is chromosomal DNA degradation. It was early on proposed that one of several known enzymes, including DNase I, DNase II, DNase γ, and cyclophilins, contribute to chromosomal DNA fragmentation during apoptosis, although the molecular mechanism understanding apoptotic DNA fragmentation by these DNases was unclear.

Most apoptotic stimuli commit the cells to apoptosis through the activation of the caspase cascade. As described in section V, several groups, including our own, have established a caspase-3-inducible cell-free system, in which not only chromosomal DNA in nuclei but also naked plasmid DNA is cleaved by caspase-3-activated cell extracts, and used it to identify a responsible factor(s), designated as CAD, for apoptotic DNA fragmentation. We purified CAD from lymphoid cells using the caspase-3-inducible cell-free system. We also noticed that an inhibiting-factor(s), designated as ICAD, is present in the extracts from non-apoptotic cells and identified ICAD as a 32-kDa protein. Molecular cloning of ICAD has revealed that long-form (331 amino acids; ICAD-L) and short-form (265 amino acids; ICAD-S) of ICAD, which are generated through alternative splicing of the same messenger RNA (Fig. 4a). Independently, other groups have purified a latent form of CAD consisting of heterodimers (DFF40 and DFF45) from human HeLa cells or an active form of CAD named CPAN from human Jurkat cells, using a system similar to ours.

Both ICAD-L and -S (ICAD) are acidic proteins with isoelectric points (pI) of around 4.5. A homology search has shown that ICAD-L has high similarity to human DFF45, suggesting that mouse ICAD-L is a counterpart of human DFF45. ICAD-L is ubiquitously expressed in a variety of tissues at the same level as ICAD-S. ICAD carry two caspase-3-cleavage sites corresponding to amino acids at 117 (DEPD^{117}) and 224 (DAVD^{224}) (Fig. 4a). Analyses of point mutations at two sites have indicated that the cleavage of ICAD at Asp117 and Asp224 by caspase-3 is crucial for the inactivation of ICAD function. Interestingly, the inhibitory activity of the ICAD with single point mutations at both sites was completely resistant to treatment with caspase-3. Among the caspases 1 to 8, caspase-7 also cleaves ICAD in an *in vitro* assay. However, active caspase-7 translocates from cytosol to mitochondrial and microsomal fractions, whereas active caspase-3 is still present in cytosol during Fas-induced apoptosis in mouse liver *in vivo,* suggesting that ICAD in cytoplasm and nuclei are mainly cleaved by caspase-3. This is consistent with the observation that caspase-3-deficient cells undergo apoptosis without chromosomal DNA fragmentation, although caspase-7 is activated.

ICAD are thermostable at 90°C for 5 min and resistant to denaturants such as 6 M guanidium hydrochloride, 8 M urea and 0.1% SDS. ICAD specifically inhibit the CAD activity, but not DNase I and DNase II activities, by binding to it. In addition, ICAD can completely inhibit nuclear apoptosis induced by extracts from Fas-activated cells in the presence of an inhibitor of caspase-3. Furthermore, overexpression of caspase-3-resistant ICAD mutants suppresses apoptotic DNA degradation induced by diverse apoptotic stimuli including death factors, growth factor starvation, anti-cancer drugs, and γ-irradiation, indicating that CAD needs to be activated to degrade chromosomal DNA in many apoptotic situations.

Molecular cloning of CAD has revealed that the mouse CAD gene encodes a basic protein of 344-amino acids and a pI of 9.7. Mouse CAD comprises 14 cysteine residues; methionine and cysteine residues at positions 1 and 2, respectively, are removed during the maturation of the protein (Fig. 4b). Human CAD is composed of 338 amino acids and is highly homologous to mouse CAD, with an identity of 75.9% between both primary structures. Eleven cysteine residues, most of which exist as reduced thiol groups, are conserved between mouse and human CAD. The C terminal region of CAD contains a stretch of basic amino acids with the features of a nuclear localization signal (Fig. 4b). Active CAD is a very unstable protein and easily aggregates under nonreducing condition, in contrast to the CAD/ICAD-L complex. The stability of CAD is enhanced by addition of reducing reagents such as dithiothreitol (DTT) and reduced glutathione, suggesting that inter- and/or intra-molecular crosslinking of free thiols in CAD may cause the aggregation and inactivation of CAD function. Some DNases are inactivated by reagents that modify free thiols, such as iodoacetamide and N-ethyl maleimide, whereas CAD activity is not inhibited by such reagents, suggesting that free thiols in CAD are not required for its

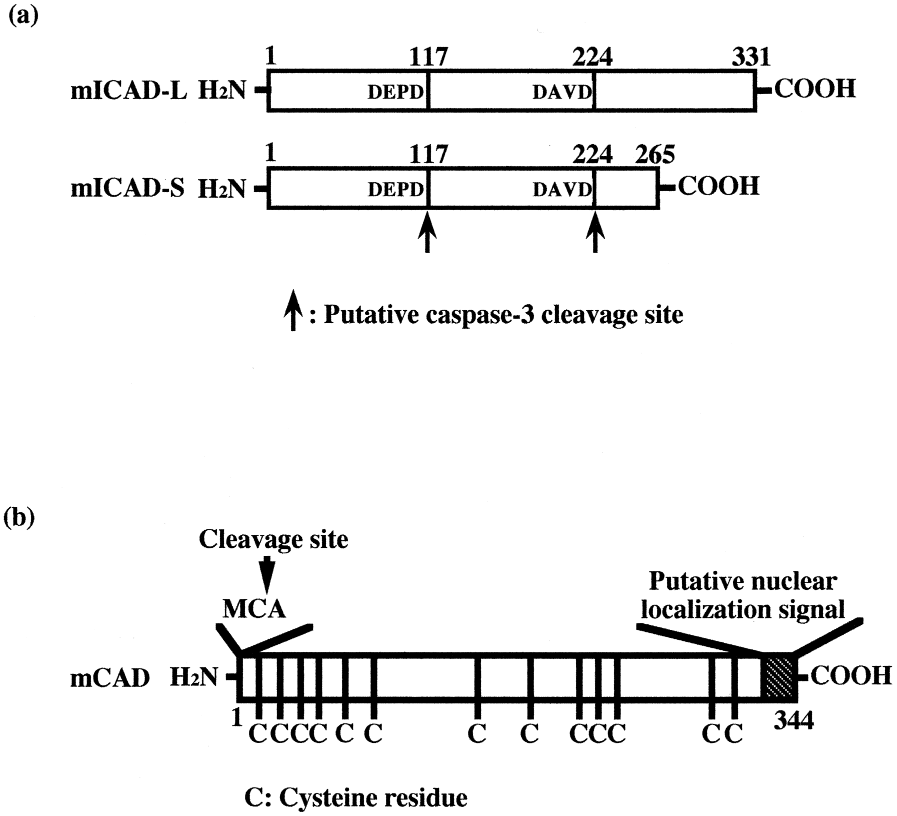

FIGURE 4 Structure of mouse CAD and ICAD.

activity. In addition, it is known that DNA increases the stability of DNase I. Therefore, it would be interesting to determine whether free-thiol-modifying reagents and/or addition of DNA could enhance CAD stability and facilitate the structural analyses.

Various endonucleases are implicated in apoptotic DNA degradation, including nonmetal ion-requiring, Mg^{2+}-requiring, Mg^{2+}- and Ca^{2+}-requiring, and Mg^{2+}- or Ca^{2+}- requiring DNases. CAD belongs to the class of Mg^{2+}-requiring endonucleases. The optimal pH for CAD activity is around 7.5. CAD activity is inhibited by Zn^{2+}, in accordance with the observation that Zn^{2+} inhibits apoptotic DNA degradation in some intact cells. CAD does not digest single-stranded DNA or RNA. It cleaves double-stranded DNA leaving 5′-phosphate and 3′-hydroxyl ends; consistent with the fact that the DNA fragments generated during apoptosis are susceptible to end labeling by deoxynucleotidyl transferase (TUNEL reaction). Furthermore, there is no consensus CAD-cleavage sequence on either genomic or plasmid DNA, but CAD appears to preferentially cleave sequences with rotational symmetry (5′-R, R, R, Y $\Downarrow$ R, Y, Y, Y-3′) or with AT-rich region. In addition, CAD cleaves chromatin better than micrococcal nuclease. So far, micrococcal nuclease has been used for chromatin research to investigate chromatin functions but CAD might be a useful tool for such analyses.

Homology search has shown that CAD has no apparent overall similarity with any proteins. It has, however, been found that CAD shares a N-terminal region consisting of about 80 amino acids, so-called CAD or cell death-inducing DFF45-like effector (CIDE) domain, with other proteins, including mouse and human ICAD, mouse Fsp27, human and mouse CIDE-A, mouse CIDE-B, and *Drosophila melanogaster* DREP-1, based on database searches. Structural analyses have shown that the CAD domains from CAD and CIDE-B have a fold of α/β roll type with a novel protein-protein-binding motif, and that ICADs and CAD appear to interact through CAD domains.

VII. MOLECULAR MECHANISM OF CAD AND ICAD

In the absence of ICAD-L, CAD is expressed in an insoluble form in various host cells including *Escherichia coli* (*E. coli*), insect cells and mammalian cells. However, coexpression of CAD and ICAD-L dramatically enhances CAD activity induced by active caspase-3, and

recombinant CAD is recovered in the cytosolic fraction as a CAD/ICAD-L complex. This process is reproduced using an *in vitro* coupled transcription and translation system. Newly synthesized CAD aggregates in the absence of ICAD-L in the *in vitro* reaction, and functional CAD is produced only in the presence of ICAD-L, although the expression level of CAD is the same whether or not ICAD-L is present. These results suggest that ICAD-L works as a chaperone to help the correct folding of CAD. These results have been confirmed by the finding that chaperone-like activity of ICAD-L is seen even in refolding of purified and denatured CAD from inclusion bodies from *E. coli* in the presence of high concentration of reducing reagents. This process does not require ATP, whereas reticulocyte lysates in combination with ICAD-L enhance a refolding process for denatured CAD in an ATP-dependent manner. These results imply that one or more ATP-dependent enhancer may participate in the folding process of CAD under physiological conditions. General chaperone systems including Hsp70 may function in this process. *In vitro* refolding studies should reveal which factor(s) is involved in the folding of CAD in the near future. Analyses of the functional differences between ICAD-L and ICAD-S have revealed that ICAD-S has less chaperone-like activity than ICAD-L and mainly exists as a homo-oligomeric complex (oligomerization of ICAD is dependent on their concentration). The finding that ICAD-L is predominantly complexed with CAD and that only ICAD-S is purified from our assay may be due to the difference of chaperone-like activity between ICAD-L and ICAD-S. Thus, when CAD is newly synthesized, ICAD-L binds to the nascent chain of CAD on the ribosome to suppress aggregation of CAD and to help proper folding (Fig. 5). ICAD-L is incorporated into a CAD/ICAD-L complex to inhibit CAD activity. Caspase-resistant ICAD-L is likely to have chaperone-like activity since the aggregation of CAD is suppressed by coexpression with it. Thus ICAD-L works as a double safeguard against dangerous CAD function. Once caspase-3 is activated by an apoptotic stimulus, ICAD-L is cleaved and released from CAD. The release of ICAD-L from complex permits active CAD to concentrate in nuclei and to degrade chromosomal DNA (Fig. 5). Mouse CAD lacking the nuclear localization signal at the C terminus, (consisting of amino acids position 3 to 329, with the 15 basic amino acids of CAD primary sequence deleted), still has DNase activity, but it cannot induce DNA fragmentation in nuclei. These observations suggest that the C terminal basic region actually works as nuclear localization signal and is not required for DNase activity.

It is thought that there are two steps in apoptotic DNA degradation. At the initial stage of apoptosis, chromosomal DNA is cleaved into 50- to 300-kilobase pair (kb)-size fragments, followed by the cleavage of large fragments to nucleosomal units. Cyclophilins have been proposed as candidates for large chromosomal degradation. However, the expression of caspase-resistant ICAD in cells blocks not only small-size nucleosomal degradation but also large-size chromosomal degradation induced by apoptotic stimuli such as Fas and staurosporine, indicating that CAD is responsible for both steps. Large-size fragments could be due to preferential cleavage at nuclear scaffolds with AT tracts by CAD. This is consistent with the fact that no large-size DNA degradation was detected in the thymocytes from ICAD-deficient mice during dexamethasone-, etoposide-, and staurosporine-induced apoptosis.

Apoptosis is accompanied by nuclear condensation as well. When active CAD is incubated with isolated nuclei, CAD itself has an ability to induce apoptotic morphological changes with chromatin condensed around the nuclear periphery. ICAD-deficient thymocytes also

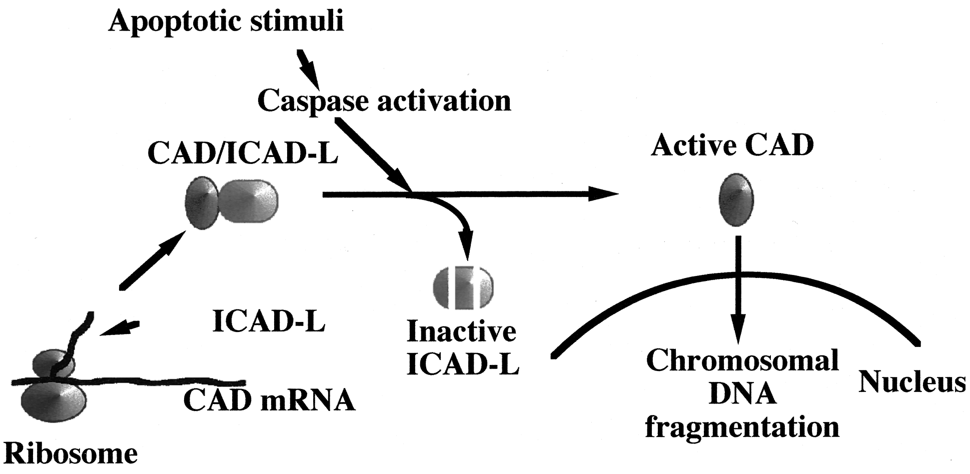

FIGURE 5 Model for the function of CAD and ICAD in apoptosis.

show no chromatin condensation in nuclei after apoptotic stimulation. On the other hand, significantly, dying cells overexpressing caspase-resistant ICAD show apoptotic chromatin condensation without DNA fragmentation. These results suggest that apoptotic chromatin condensation is caused by CAD in particular tissues such as thymocytes and that another factor(s) is involved in this process in some situations. We also cannot rule out the possibility that CAD has dual functions, DNA fragmentation and condensation activities, because of complete denatured structure of CAD in ICAD-null cells (CAD is present as an insoluble form) but not transformants expressing caspase-resistant ICAD (endogenous CAD and ICAD are still present as a soluble form), or that cleaved ICAD cooperatively work with cellular factor(s) to condense chromatin although cleaved ICAD themselves have no chromatin condensation activity. Recently, it has been reported that a protein other than CAD, called Acinus, is responsible for apoptotic chromatin condensation however this remains to be confirmed.

VIII. PHYSIOLOGICAL DNA FRAGMENTATION AND PHAGOCYTOSIS OF APOPTOTIC CELLS

Extensive DNA fragmentation and chromatin condensation are hallmarks of apoptosis and are tightly regulated by the CAD/ICAD system. It is thought that these processes play an important role in cell killing. To reveal physiological functions of the CAD/ICAD system *in vivo,* Zhang *et al*. have established ICAD-deficient mice. These mice develop normally compared to wild-type mice and do not show any pathological signs. However, thymocytes from ICAD-null mice are obviously lacking CAD activity and show neither DNA fragmentation nor chromatin condensation following exposure to apoptotic stimuli. Consistent with the observations described above, no functional CAD is expressed, at least in thymocytes, in the absence of ICAD. Moreover, transgenic mice expressing caspase-resistant ICAD-S (ICAD-Sdm; it is ubiquitously expressed under the control of elongation factor 1 α promoter) have been established. ICAD-Sdm is expressed in the major organs of the transgenic mice, and isolated thymocytes from the transgenic mice undergo apoptosis without DNA fragmentation after irradiation with γ-rays. However, these transgenic mice have no phenotypic abnormality during development, like ICAD-deficient mice. The thymocytes from ICAD-null mice, but not from wild-type mice, are partially resistant to several apoptotic stimuli including dexamethasone, etoposide, and staurosporine, whereas thymocytes from ICAD-Sdm transgenic mice die as efficient by as those from wild-type mice following exposure to γ-rays. Why is there such a discrepancy even though there is no CAD function in thymocytes in both cases? There is a possibility that either endogeneous ICAD cleaved by caspase-3 or CAD/ICAD-Sdm complex may contribute to some killing activity in thymocytes from ICAD-Sdm transgenic mice. ICAD might be required for amplification of death signal in some conditions, or the modification of anti-apoptotic proteins in cells. It has recently been reported that ICAD-null mice exhibit enhanced spatial learning and memory accompanied by the increase ($\sim$8%) of granule cells in the hippocampal dentate gyrus region compared to wild-type mice. Detail analyses would be elucidated whether or not CAD/ICAD system is actually involved in the induction of cell death.

It has been thought that phagocytes recognize the apoptotic cells undergoing DNA fragmentation and ingest and recycle the dead cells, and that cleaved DNA is utilized as a marker for the recognition of apoptotic cells by phagocytes. However, phagocytosis of dying thymocytes from ICAD-Sdm transgenic mice, in which DNA degradation is not seen, occurs as efficiently in the case of wild-type mice, indicating that DNA fragmentation is dispensable for phagocytosis. Moreover, it has recently been observed that TUNEL-positive cells are still present in various tissues from ICAD-Sdm transgenic mice at similar levels as in wild-type mice after γ-irradiation. When dying cells overexpressing ICAD-Sdm are cocultured with macrophages, their intact DNAs are quickly processed into nucleosomal units. Furthermore, reagents that block the acidification of lysosomes inhibit DNA degradation induced by macrophages, suggesting that lysosomal DNase in engulfing cells is involved in the generation of TUNEL-positive cells in various tissues *in vivo*. A candidate for the responsible DNase during phagocytic DNA fragmentation is DNase II existing in lysosomes as an acidic DNase. The recent molecular cloning of *nuc*-1, which is responsible for DNA degradation during programmed cell death in *C. elegans*, has revealed that the gene product is a homologue of mammalian DNase II. This report proposes a model for apoptotic DNA degradation composed of three distinct steps (Fig. 6). The first step appears to involve an unidentified endonuclease to produce TUNEL-positive DNA. The enzyme working in this step would be a CAD-like DNase. In the second step NUC-1 converts TUNEL-positive to TUNEL-negative DNA. In the final step, “cell-corpse DNA” is completely digested by engulfment-dependent nuclease(s). DNase II creates DNA fragments having 3′-phosphate ends rather than the 3′-hydroxyl ends that are required as primers for terminal deoxynucleotidyl transferase. In the mammalian system, after cleavage of DNA by lysosomal DNase II, the 3′-phosphate groups are presumably removed by acid

phosphatase, which is abundant in lysosomes, giving rise to TUNEL-reactive ends. Figure 6 summarizes the model proposed for apoptotic DNA degradation in the mammalian system, based on the *C. elegans* paradigm: Once caspase-3 is activated by one of various apoptotic stimuli, ICAD is inactivated and releases from CAD. Activated CAD cleaves chromosomal DNA generating TUNEL-reactive ends. This is followed by DNase II-mediated DNA degradation that gives rise to TUNEL-negative DNA (Fig. 6). Due to participates of lysosomal enzymes containing a large amount of acid phosphatase in mammalian phagocytes, TUNEL-reactive ends could be generated in engulfing cells in mammals. Finally, complete digestion of cell-corpse DNA occurs during phagocytosis to release and recycle free nucleotides (Fig. 6). Recently, a CAD/ICAD system has been identified in *Drosophila melanogaster*. Thus, the mechanism for apoptotic DNA degradation is likely to be conserved between mammals, flies, and worms. The detailed mechanism will be elucidated by further use of genetic and molecular biological approaches.

A variety of factors contributing to engulfment have been reported in nematodes, flies and mammals. Some of the responsible factors, including CED-1, CED-2, CED-5, CED-6, CED-7, CED-10, and CED-12 in *C. elegans*, Croquemort (CRQ) in *Drosophila melanogaster*, and a phosphatidylserine receptor in mammals have recently been identified. The CED-6 protein contains a phosphotyrosine-binding (PTB) domain at its N terminus and putative Src-homology domain 3 (SH3)-binding sites. It has been shown that this protein restores the function of the clearance of apoptotic cells in engulfment-defective *ced*-6 mutants from *C. elegans*. The CED-7 protein is homologous to ABC (ATP-binding cassette) transporters, is localized in the plasma membrane and the endoplasmic reticulum, and appears to function both in dying cells and in engulfing cells during the engulfment process. It has been suggested that CED-7/ABC1 is required for the translocation of molecules that interact between the cell surfaces of the apoptotic cells and phagocytes in both *C. elegans* and mammals. The *ced*-5 gene product, which functions in cell-corpse engulfment in *C. elegans*, is similar to the human DOCK180 and the *Drosophila melanogaster* Myoblast City (MBC). It has been suggested that both DOCK180 and MBC are involved in the extension of cell surfaces and may function during phagocytosis in different species. CED-2 and CED-10 are counterparts of the human adapter protein CrkII and the human GTPase Rac, respectively. These factors are likely to function in phagocytes, but not in dying cells, to regulate phagocytosis. Based on genetic crossing experiments, CED-10 appears to be downstream of CED-2 and CED-5 and GTPase signaling must regulate the polarized extension of cell surface for phagocytosis. *Drosophila* CRQ is a CD36-like receptor expressed on the cell surface of macrophages; genetic analyses have shown that it is required for phagocytosis of apoptotic corpses. Apoptosis is also accompanied by the exposure of phosphatidylserine on the outer cell surface of plasma membranes. Phosphatidylserine has also been regarded as a target molecule for phagocytosis; the phosphatidylserine receptor has been cloned by phage display technique. It is expressed on various cells including macrophages, fibroblasts, and epithelial cells, and is found in databases of *Drosophila melanogaster* and *C. elegans* as a protein of unknown function. An antagonistic antibody against the phosphatidylserine receptor blocks engulfment of dying B and T cells by phagocytes, indicating that this receptor is essential for phagocytosis of

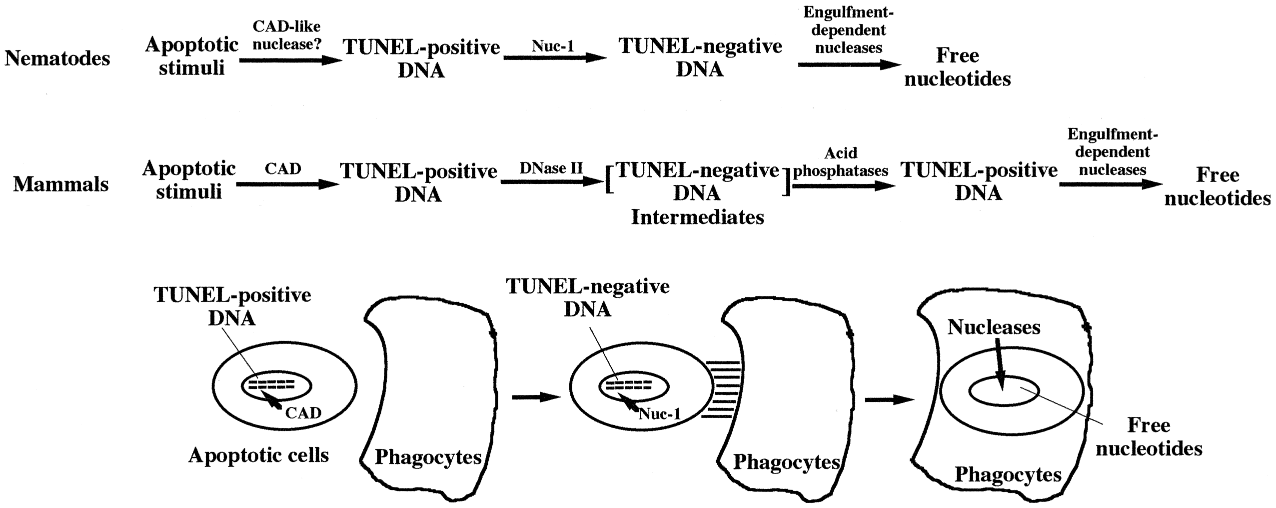

FIGURE 6 Model for apoptotic chromosomal DNA degradation *in vivo*.

at least some apoptotic cells. Thus, although various factors participating in phagocytosis have been identified, it will be necessary to find further components involved in this process to understand the molecular mechanism of clearance of apoptotic cells in our bodies.

IX. PERSPECTIVES

Not only cell growth and differentiation factors but also death factors regulate homeostasis of organisms. The death factors activate a caspase cascade through their specific receptors. Although it is said that procaspase-8 is self-cleaved simply following activation by oligomerization, the precise mechanism is still unclear. Some unknown additional factor(s) may be required for this process to generate active caspase-8 more efficiently and remain to be identified by future analyses. Furthermore, it is known that there is caspase-independent apoptosis. It would be important to investigate what kinds of factors are involved in such apoptotic pathways. The mechanism of apoptotic DNA fragmentation has been largely clarified by biochemical approaches, but the mitochondrial pathway for death signals remains elusive. How are apoptotic signals transduced in mitochondria? Which factors participate in mitochondrial apoptosis? Cell-free systems using isolated mitochondria may lead to a satisfactory solution of these problems.

Cell death occurs without cell-autonomous DNA degradation in cells devoid of CAD activity and developmental processes appear to be normal (without drastic changes) despite the lack of functional CAD. Why has a complicated CAD/ICAD system evolved in our bodies? One possibility is that this system precludes the transformation of phagocytes by dangerous genes such as oncogenes or viral genes. Another possibility is that the cell-autonomous DNA fragmentation mechanism reduces autoimmune responses against nucleosomal DNA, which is known as a strong autoantigen. Nucleosomal DNA released from apoptotic cells during developmental processes could lead to autoimmunity. Alternatively, the CAD/ICAD system may promote the recycling of the DNA components. Analyses of CAD/ICAD system-deficient mice may cast light on these questions. Furthermore, the number of genes recognized as being involved in phagocytosis of apoptotic cells is gradually increasing due to genetic and molecular biological analyses. Thus, the understanding of detail molecular mechanisms of apoptosis deepens as science is progressed, so the elucidation of overall mechanisms for apoptosis through death receptors may not be so far off.

ACKNOWLEDGMENT

I thank all members of Dr. Nagata's laboratory in Osaka Bioscience Institute and Osaka University Medical School for helping during my initial works. I am grateful to Drs. S. Nagata and H. Sakahira in Osaka University Medical School for helpful comments on the manuscript and to Dr. C. Weissmann at the Imperial College School of Medicine for supports, discussion, and encouragement.

SEE ALSO THE FOLLOWING ARTICLES

Chromatin Structure and Modification • Gene Expression, Regulation of • Immunology-Autoimmunity • Mammalian Cell Culture • Ribozymes • Translation of RNA to Protein

BIBLIOGRAPHY

Budihardjo, I., Oliver, H., Lutter, M., Luo, X., and Wang, X. (1999). "Biochemical pathways of caspase activation during apoptosis," *Annu. Rev. Cell Dev. Biol.* **15,** 269–290.

Bratton, S. B., MacFarlane, M., Cain, K., and Cohen, G. M. (2000). "Protein complexes activate distinct caspase cascades in death receptor and stress-induced apoptosis," *Exp. Cell Res.* **256,** 27–33.

Ferri, K. F., and Kroemer, G. (2000). "Control of apoptotic DNA degradation," *Nat. Cell Biol.* **2,** E63-4.

Green, D. R., and Reed, J. C. (1998). "Mitochondria and apoptosis," *Science* **281,** 1309–1312.

Liu, Q. A., and Hengartner, M. O. (1999). "The molecular mechanism of programmed cell death in *C*," *Elegans. Ann. N. Y. Acad. Sci.* **887,** 92–104.

Nagata, S. (1997). "Apoptosis by death factor," *Cell* **88,** 355–365.

Nagata, S. (2000). "Steering anti-cancer drugs away from the TRAIL," *Nat. Med.* **6,** 502–503.

Nagata, S. (2000). "Apoptotic DNA fragmentation," *Exp. Cell Res.* **256,** 12–18.

Raff, M. (1998). "Cell suicide for beginners," *Nature* **396,** 119–122.

Stroh, C., and Schulze-Osthoff, K. (1998). "Death by a thousand cuts: An ever-increasing list of caspase substrates," *Cell Death Diff.* **5,** 997–1000.

Thornberry, N. A., and Lazebnik, Y. (1998). "Caspases: Enemies within," *Science* **281,** 1312–1316.

Yeh, W. C., Hakem, R., Woo, M., and Mak, T. W. (1999). "Gene targeting in the analysis of mammalian apoptosis and TNF receptor superfamily signaling," *Immunol. Rev.* **169,** 283–302.

Wyllie, A. H., Kerr, J. F. R., and Currie, A. R. (1980). "Cell death: The significance of apoptosis," *Int. Rev. Cytol.* **68,** 251–306.

Wyllie, A. H. (1980). "Glucocorticoid-induced thymocyte apoptosis is associated with endogenous endonuclease activation," *Nature* **284,** 555–556.

Zhang, J., Liu, X., Scherer, D. C., van Kaer, L., Wang, X., and Xu, M. (1998). "Resistance to DNA fragmentation and chromatin condensation in mice lacking the DNA fragmentation factor 45," *Proc. Natl. Acad. Sci. U.S.A.* **95,** 12480–12485.

Chromatin Structure and Modification

Fyodor D. Urnov
Alan P. Wolffe
Sangamo Biosciences

I. The 6-Billion Challenge
II. Chromatin: A Brief History of Scholarship
III. Chromatin Structure: The Histones and the Nucleosome
IV. The Disruption and Modification of Chromatin Structure as a Tool to Control the Genome
V. Chromatin and Transcription: A Synthesis

GLOSSARY

Chromatin A complex between genomic DNA and proteins (mostly histones) that compacts the genome into the eukaryotic nucleus and enables its functionality.

DNA methylation The covalent modification of cytosine to yield 5-methylcytosine. In vertebrates occurs exclusively within the context of a CpG dinucleotide. Produced by DNA methyltransferase and interpreted by dedicated DNA-binding proteins that effect transcriptional repression in part via recruiting histone deacetylase.

HAT Histone acetyltransferase. Enzyme that catalyzes the acetylation of the ε-NH_2 group in the side chain of lysine residues in the core histones' amino terminal tail. Commonly used in transcriptional activation.

HDAC Histone deacetylase. Enzyme that catalyzes the removal of the acetyl residue from ε-acetyllysine in the core histones' amino terminal tail. Commonly used in transcriptional repression.

Histone Small (ca. 110 amino acids), highly basic protein that is rich in lysine and arginine. Assumes a distinctive 3-helix "histone fold" tertiary structure with an extended amino terminal tail. Four "core" histones (H2A, H2B, H3, H4) form an octamer that winds DNA onto itself to form the nucleosome core particle.

Nuclear hormone receptor (NHR) A transcription factor whose activity is regulated by a small molecule (the ligand), such as a steroid (estrogen, cortisol, etc.) or an amino acid (thyroid hormone). Most NHRs are transcriptional activators in the presence of ligand; some NHRs also act as transcriptional repressors in its absence.

Locus (pl. *loci*) Geneticists' term for "specific location on the chromosome (in the genome)."

Nucleosome A union of eight core histones, one linker histone, and ca. 160 base pairs of DNA. The elementary building block of the chromatin fiber.

Remodeling An ATP-dependent process of disrupting histone–DNA contacts in chromatin. Requires the action of dedicated macromolecular machines such as the SWI/SNF complex. An integral component of gene activation and repression.

Transcription The process of RNA synthesis on a DNA template. Catalyzed by RNA polymerase.

COMPACTION of the eukaryotic genome into the nucleus must occur such that the instructions in the DNA are accessible to molecular machines that effect replication, transcription, and repair. This challenging task is performed by uniting the DNA with histone proteins to yield chromatin—a dynamic, complex structure that controls the genome by both compacting and revealing it. Each 180 base pairs of DNA is complexed with eight molecules of core histone and one linker histone to yield the nucleosome. A wide variety of modifications and alterations of the nucleosomal fiber occur *in vivo*, including ATP-dependent disruption of histone-DNA contacts, and the covalent modification of the histone tails by acetylation. The enzymatic machines that generate such modifications are targeted to specific loci in the genome by various regulators to effect gene activation and repression.

The enormously complex program of gene expression that unfolds during ontogeny in eukaryotes is unequivocally contingent on an accurate interpretation of the genome's contents by transcriptional regulators. Their action is abetted by, and intimately functionally linked to, the native structure assumed by their DNA target *in vivo*: the nucleoprotein fiber of chromatin. This complex union of DNA with histone and other proteins ensures that the genome is accommodated within the nucleus by winding each 180 base pairs of DNA into a nucleosome, and by subsequent higher-order folding of the nucleosomal fiber. Chromatin is not a passive scaffold, and undergoes a wide variety of structural transitions in response to developmental, environmental, and internal gene regulatory stimuli. A host of dedicated enzymatic complexes effect these transitions both in a genome-wide (i.e., after DNA replication) and a localized fashion, when they are recruited by transcriptional regulators to mediate gene actvation or repression at specific loci. Unique patterns of covalent post-translational histone modification, and perturbation of histone-DNA contacts or higher-order chromatin fiber structure that result from such targeting exert a potent regulatory effect on the underlying DNA.

I. THE 6-BILLION CHALLENGE

The functionality of the human genome—i.e., its capacity to control an extraordinarily complex program of cell division, differentiation, and morphogenesis during embryonic development and adult ontogeny—is enabled in the face of several formidable challenges.

The first one is topological: the combined length of a single cell's worth of human DNA is ca. 2.3 m (there are $\sim 6.6 \times 10^9$ base pairs of DNA in the diploid human genome and each base pair is 0.32 nm long) and this extraordinary quantity of nucleic acid—in the form of 46 separate molecules—is packaged into that cell's nucleus, i.e., a sphere with a diameter of $\sim 5\ \mu$M. Thus, the length of the DNA fiber (which is 2 nm wide) exceeds that of its natural container by a factor of ∼500,000.

The second challenge is that of the signal-to-noise ratio: laden with molecular atavisms and genomic parasites, the genome's ∼800 Mb of genetic information (counting each base pair as two bits, i.e., four base pairs per byte) make *Finnegan's Wake* seem lucid and succinct by comparison, since only ca. 20 Mb (∼3%) represent DNA that codes for protein. An even smaller fraction is "regulatory DNA"—i.e., stretches of the genome that contain molecular instructions on how and when its ∼40,000 genes are to be activated and silenced. The exact percentage of the genome assigned to such regulatory function is unknown, but a conservative estimate would be ca. 1% (8 Mb). Thus, the genome is a remarkably messy manual, in which a deluge of genetic gibberish (or, in less emphatic terms, of DNA to which we have not yet been able to assign a function) hides the occasional useful instruction or snippet of data. The molecular complexes that have evolved to interpret the genome—sequence-specific DNA binding proteins and their cofactors—are thus challenged by a major problem of parsing the content of the genome in search of relevant information.

As if this were not enough, the physicochemical properties of DNA—a very long polymer—make its aqueous solutions extraordinarily viscous (as anyone who has ever worked with human genomic DNA knows from experience, solutions of even 10 mg/ml cannot be pipetted if the DNA is not sheared first into fragments of smaller size). All interactions relevant to the biology of the genome, however, occur—and quite robustly—in an aqueous environment that contains ∼100 mg/ml DNA (6–8 pg in ca. 65 femtoliters)!

All these challenges are very successfully met: inside the nucleus, the DNA of the genome finds itself a powerful ally as a large number of dedicated molecular machines assemble it into a highly structured fiber by winding it around specialized proteins called "histones";

the resulting entity—chromatin—presents the genome in compact, interpretable, and soluble form, and—most importantly—enables a remarkably rapid response to a great variety of internal and external stimuli.

II. CHROMATIN: A BRIEF HISTORY OF SCHOLARSHIP

Cytological, genetic, biochemical, and biophysical studies of the past 100 years have offered many insights into the structure of chromatin and chromosomes and into the mechanistic aspects of its role in enabling gene expression and chromosome behavior. Our current notions of chromatin structure and function offer interesting testimony to the validity of T. Kuhn's well-known thesis on the role of "paradigms" in scientific inquiry. The first half of the 20th century witnessed the protein constituents of the nucleus being awarded the role of carriers of genetic information, with DNA relegated to a minor role of a scaffold. After the discovery of the nucleosome in 1973–1974, some 20 years of scholarship in eukaryotic transcription proceeded under the auspices of notions developed in studying gene regulation in bacteria, and thus the histones underwent quite the proverbial reversal of fortune and were considered a repressive, obstructive scaffold instead; *bona fide* transcriptional control was only thought to occur on stretches of naked, histone-free DNA.

Following the genetic and biochemical characterization of complex machines that modify chromatin in the mid-1990s, the past few years have offered a dramatic and emphatic shift in our appreciations of its intranuclear function: no longer viewed as monotonous or merely repressive, chromatin is now considered an essential component of most gene regulatory pathways *in vivo*. We now see the nucleus as populated with enzymatic complexes that remodel chromatin in a targeted fashion to achieve a wide variety of regulatory responses from the DNA. The current view of transcriptional and genomic control, therefore, is one of a complex and mutually beneficial symbiosis between the protein regulators of the genome and the nucleoprotein architecture of chromatin.

The etymology of the term "chromatin" traces its origins to cytological studies of the late 19th century, when "thread-like structures"—chromosomes—were revealed in dividing cells by staining with dyes. These were presumed to consist of "chromatin" (a term proposed by W. Waldeyer in 1888) whose chemical nature was completely obscure. Fortunately for science, the contemporaneous discovery by F. Miescher of DNA in human lymphocytes eventually prompted an analysis into whether the substance of chromosomes is in any way related to the chemical entity extracted by Miescher from cells known as "nuclein." In 1889, work by R. Altmann demonstrated that it consists of nucleic acid and protein. A functional connection between either of these two chemical substances and the phenomenon of heredity remained elusive, however.

In the 1900–1920s, studies by T. H. Morgan and a distinguished pleiade of his junior colleagues (C. B. Bridges, H. K. Muller, and A. H. Sturtevant) in the fruit fly *Drosophila melanogaster* physically placed genes onto chromosomes, connecting genetics and cytology. Biochemical analysis performed by A. Kossel in the 1890s–1910s led to the discovery of histones and protamines—the small, highly positively charged protein components of chromosomes. Their apparent biochemical diversity, coupled with a notion of DNA as a chemically monotonous entity, prompted the belief into their role as carriers of genetic information. This theory was placed in very strong doubt in 1944 by work of O. T. Avery, C. M. MacLeod, and M. McCarty at the Rockefeller Institute, and firmly laid to rest in 1952 by A. D. Hershey and M. Chase at the Cold Spring Harbor Laboratory.

After a combined effort by chemists and X-ray crystallographers, including P. Levene, A. Todd, E. Chargaff, R. E. Franklin, and M. H. F. Wilkins, made their indelible contribution to the 1953 discovery by J. D. Watson and F. H. C. Crick that DNA is a double helix, a 20-year avalanche of experimentation revealed the foundation of genetic information maintenance and transfer in living systems, mostly prokaryotes. This period witnessed three very important studies that implicated chromatin in effecting eukaryotic gene regulation.

Cytological studies of mammalian cells have revealed an interesting gender bias: the interphase (i.e., nondividing) nuclei of female, but not male, cells contained a small, microscopically dense entity termed the "Barr body." In 1958, work from S. Ohno demonstrated that this entity corresponds to one of the two X chromosomes in the female karyotype. Soon after, in 1961, genetic experiments by M. Lyon showed that the female genome is functionally hemizygous for sex-linked loci—i.e., that of the two X chromosomes in mammalian females, one is genetically silent (this phenomenon is known as "X chromosome inactivation"). The combination of these data provided evidence for a correlation between the structure of specific chromosomes and their state of activity. In recent years, work from many laboratories, most notably those of S. Tilghman and R. Jaenisch, offered several remarkable insights into the mechanistic aspects of X chromosome inactivation (see below). Importantly, observations made by Ohno and Lyon extended earlier (1928) studies by E. Heitz on plant chromosomes, which lead to the discovery that the contents of the nucleus can be divided into heterochromatin and euchromatin,

i.e., microscopically distinct compartments that appeared to represent chromosomes that condensed to distinct degrees: heterochromatic domains in chromosomes were thus correctly deduced to contain silenced genes.

The complementary correlation between chromosome decondensation and gene activation was provided by studies in insects, whose salivary glands contain giant chromosomes that yield easily to microscopic examination. A terminally differentiated tissue destined for destruction in the metamorphosis from a larva to the adult insect, the salivary gland effects a very interesting solution to a demand for high protein production: it replicates its DNA without intervening mitosis, and all the resulting DNA fibers for each chromosome coalesce—side by side—into a macroscopic body termed "the polytene chromosome." Their size and certain other structural features have made them invaluable model systems in biology. Most importantly, cytologic analysis of changes in chromosome structure during larval development have revealed that at defined timepoints, specific stretches of these giant chromosomes decondense and form "puffs"—localized swellings. It was correctly deduced quite early on that such decondensation is related to the activation of genes that reside in those stretches of the chromosome.

In 1960, H. Clever and U. Karlsson combined their efforts of trying to determine, why the insect molting hormone, the steroid ecdysone, causes such dramatic morphological changes during insect development. The injection of purified ecdysone into larvae of the midge *Chironomus* lead to a premature and dramatic puffing of specific stretches of polytene chromosomes: thus, hormone action was for the first time connected to regulation of gene activity and, importantly, to a localized alteration (decondensation) of chromosome structure. A molecular mechanism, or correlate, of this striking phenomenon was not obtained until studies in 1984 by K. Zaret and K. Yamamoto, and subsequent experiments by T. Archer and G. Hager (Section IV.A).

Finally, 1964 saw the publication of a discovery whose impact resonated only in 1996, but quite emphatically: the observation by V. Allfrey, and A. E. Mirsky that histone proteins are subjected to postranslational covalent modification via the acetylation and methylation of lysine residues in their NH_2-terminal tails (Sections IV.A and IV.C). Because the modifications reduce the positive charge of the histones (and thus have the potential to alter the way histones interact with DNA), it was immediately suspected they might have regulatory consequences. Conclusive evidence to that effect was obtained in 1998 (Section IV.C).

Inherent technical limitations of cytological and biochemical methods described could not illuminate the molecular structure of protein–DNA contacts within chromatin. The application of biophysical techniques in the early 1970s led to significant progress in this field, and structural understanding improved accordingly, as a 7Å X-ray crystal structure of the elementary subunit of chromatin—the "nucleosome core particle" (146 bp of DNA wrapped around the an octamer of core histones)—was described in 1984, and a 3.1Å resolution structure—in 1991. Finally, in 1997, T. J. Richmond and colleagues published a 2.8Å nucleosome core particle structure—thus, we now understand the elementary composition and structure of chromatin in very considerable detail. The next section describes the histone proteins, genetic evidence for their role in regulating transcription, work in the 1970s that led to the nucleosome hypothesis, and concludes with a description of the structure of the nucleosome.

III. CHROMATIN STRUCTURE: THE HISTONES AND THE NUCLEOSOME

A. The Histones

As discovered by Kossel early this century, the primary protein residents of the nucleus are small highly basic proteins called "the histones" (there are approximately 36 million nucleosomes, each containing nine histone molecules and 180 bp of DNA, in the nucleus of a human cell). Only five different histones are sufficient to assemble chromatin: four core histones (H2A, H2B, H3, and H4; two of each are present in each nucleosome) and one linker histone (H5; one per nucleosome). A technological Atlas for molecular biology, gel electrophoresis through such matrices as polyacrylamide and agarose has been an invaluable tool in studying the genome, and Fig. 1 shows a denaturing (i.e., performed in the presence of detergent, such as sodium dodecyl sulphate, SDS) polyacrylamide gel on which the histones are resolved.

The histones' primary amino acid sequence offer several glimpses into their function: these proteins are small (between 102 and 130 amino acids long) and very rich in lysine or arginine. For example, of the 103 amino acids in human histone H4, 11 are lysine and 14 are arginine. This has an immediate electrostatic implication for histone behaviour *in vivo*: the pK_a values for these amino acids' side chains (10.0 and 12.0, respectively) indicate that at physiological pH, the average histone H4 molecule carries ca. 24.99 positive charges on itself.

An interesting and informative aspect of histone biology is the extraordinary degree of sequence conservation between these proteins across taxa: histone H4 in humans and in tomato (*Lycopersicon esculentum*) is identical in length and sequence with the exception of three highly conservative substitutions (i.e., $Val_{61} \rightarrow$ Ile). Thus, there

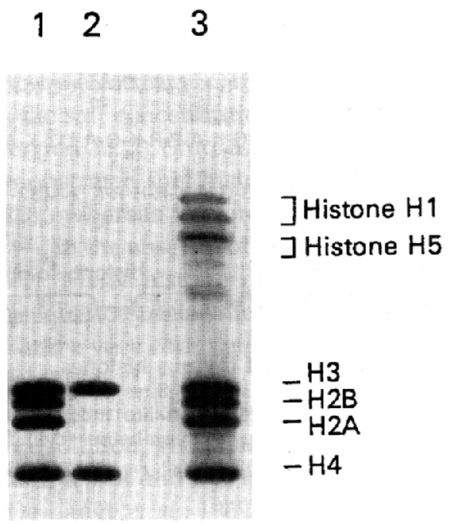

FIGURE 1 The core histones (lane 1), histones H3 and H4, or the core histones together with the linker histones (lane 3) resolved on a polyacrylamide gel.

is considerable irony in the brief (ca. 1900–1952) historical prominence that histones played as the putative vehicles of genetic data: DNA was incorrectly throught to be monotonous in sequence, but histones far exceed the DNA in their invariance, both from nucleosome to nucleosome, and between species. Such remarkable resistance to mutational pressure is particularly striking when one considers that—as discussed in the next section—only 75% of each histone is actually required to assemble a nucleosome core particle. As shown in Fig. 2, all four core histones can be assigned an identical secondary structure: the COOH-terminal 75% wind into 3 α-helices separated by two loops (this portion of the histone functions in assembling the histone octamer and in arranging DNA onto itself), while the structure of the NH_2-terminal 25% ("the tail") is not known. The conservation of the primary amino acid sequence of the tail is strong evidence for its functional prominence, which is discussed below.

The "linker histone"—histone H1—is so named because of its physical location in the chromatin fiber (see following). It is slightly larger than the core histones (ca. 20,000 Da), very rich in lysine, and assumes a very interesting structure shown in Fig. 3, called "the winged helix" (helix 3 is thought to contact the DNA). Several transcriptional regulators in metazoa are isomorphous to histone H1, and the implications of this near-congruence for their physical location within chromatin and mechanisms whereby they affect gene expression are discussed here.

B. Experimental Evidence for the Histones' Role in Controlling Gene Expression

Before a discussion of chromatin structure details, it is helpful to summarize *in vivo* evidence indicating that histones play a role in gene regulation. Simple *a priori* considerations suffice to realize that histones are required for cell viability—since chromosome condensation and subsequent segregation during mitosis is contingent on histones, cells would fail to divide in their absence and die, and this complicates genetic analysis (as eloquently put by one molecular biologist, "dead cells don't tell stories"). An elegant solution was implemented by M. Grunstein, in whose lab a strain of budding yeast, *Saccharomyces cerevisiae*, was engineered for such a study: the promoter of the histone H4 gene was replaced with one that could be inactivated with a simple change in growth medium (in this case, the addition of glucose). The gradual depletion of histone H4 from chromatin does not lead to instant cell lethality, and thus effects of "genetically dechromatinizing DNA" can be studied.

Use of such approaches by Grunstein's lab, as well as of M. Osley, F. Winston, and M. Smith, offered a very unexpected result: general notions of chromatin as repressive packing material would indicate that most of the genome should become spuriously active in the absence of chromatin. Experimental observations indicated

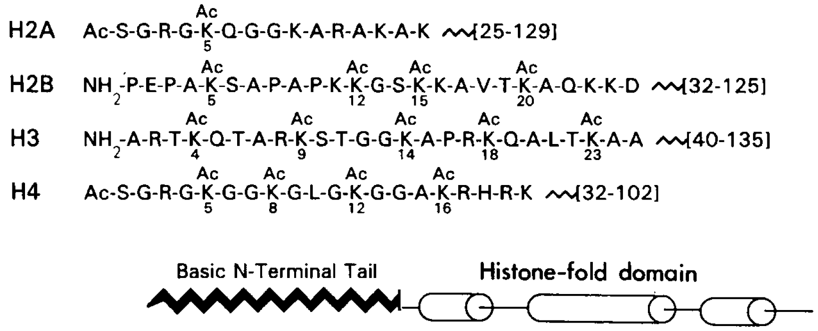

FIGURE 2 The core histones. The sequence of the amino terminal tails is indicated. The COOH-terminal domain is depicted schematically at the bottom.

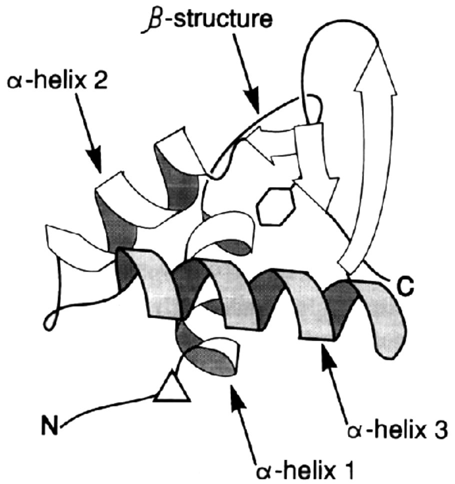

FIGURE 3 The linker histone—a ribbon representation of its "winged-helix" structure.

otherwise, however, and showed that histone depletion—and a concomitant decrease in the extent to which DNA is assembled into chromatin—does not have nucleus-wide transcriptional consequences, although the upregulation of specific genes was indeed observed.

Conclusive evidence to this effect was recently provided in work from the lab of R. Young. This study made use of the fact that the entire genome of budding yeast has been sequenced, and it was therefore possible to perform genome-wide expression profiling analysis. This remarkable experiment requires a custom-made "microarray": a silicon chip containing a grid, into each cell of which a different single-stranded nucleic acid probe corresponding to a given yeast gene is placed and immobilized (a total of 5900 genes were thus analyzed in one experiment). Messenger RNA is prepared from wild-type and mutant cells at defined timepoints following inactivation of the histone H4 gene, and the change in the levels of each message is determined by hybridizing each mRNA sample to a separate chip (in actual fact, for detection purposes, a copy of the mRNA labeled with a fluorescent dye is prepared) and then measuring the difference—if any—in the levels of the mRNA hybridizing to each cell in the array.

The major result from this study is that expression of ca. 75% of the yeast genes is not significantly altered by histone depletion. This indicates that in budding yeast, chromatin does not have a genome-wide transcriptional repression function. Remarkably, of the remaining 25% of the genes, ca. 15% were upregulated more than threefold, while 10% were *downregulated* more than threefold. These data suggest that nucleosomes have *gene-specific* roles in transcriptional control, and, surprisingly, that the assembly into chromatin is not only required for the repression of some genes but also for the activation of some others.

As elaborated in the next section, a examination of the nucleosome structure fails to explain these data—all nucleosomes *appear* to be the same by available structural criteria, and assemble the entire genome into what *seems* to be a homogeneous fiber (see following). Remarkably, thinning out that fiber has markedly idiosyncratic effects on genome behavior—and some understanding as to possible mechanisms for enabling such nonuniform responses is beginning to emerge.

C. Ontology of the Nucleosome

Evidence that DNA in the nucleus may be organized into some sort of repetitive entity first came from analysis of the way the genome inside the nucleus is seen by nucleases (i.e., enzymes that degrade DNA by cleaving the phosphodiester bond between two adjacent nucleotides), both endogenous to the cell and ectopic. Such experiments by R. Williamson in 1970, and by D. Hewish and L. Burgoyne in 1973, demonstrated that DNA released from the genome in this way assumes a nonrandom length distribution—which was unexpected, because nucleases were not known to have a substrate preference within DNA. By way of example, Fig. 4 shows an agarose gel containing DNA samples after treatment with a nuclease. When the enzyme

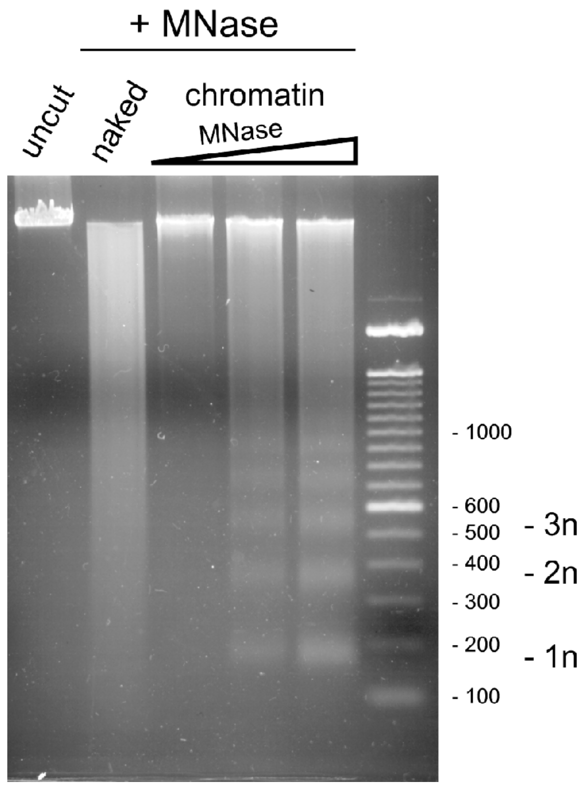

FIGURE 4 Chromatin treatment with micrococcal nuclease (MNase) yields a nonrandom distribution of DNA fragments—evidence for the nucleosome's existence.

encounters DNA in naked form, random degradation of the phosphodiester backbone occurs, and a relatively homogeneous distribution of DNA sizes is visualized (if the reaction were allowed to proceed longer, DNA would be eventually degraded to mononucleotides). In contrast, when *cell nuclei* are treated with the same nuclease, rather than degrade the genome in an identical manner, the enzyme generates populations of discretely sized DNA fragments that appear to occur in multiples of 180 base pairs. This suggest that *in vivo*, the DNA is somehow packaged into "180 base pair installments" such that only the DNA stretch between two adjacent 180 base pair "packages" is accessible to the nuclease. At the same time, analysis of the hydrodynamic properties of individual DNA–protein particles released by nuclease using analytical centrifugations, allowed K. van Holde and coworkers to measure its molecular weight at ca. 180,000 Da.

Very strong support for the notion of chromatin being composed of a reiteration of identical subunits came from electron microscopic studies by C. Woodcock and other scientists in 1973–1974. When preparations of chromatin were spread under appropriate ionic conditions on a carbon grid and visualized under the EM, a remarkable "bead-on-a-string" fiber was visualized (Fig. 5). Cross-linking experiments by J. Thomas and R. Kornberg indicated that the histones' representation in chromatin is stoichiometric; the combined weight of all these data led to R. Kornberg's proposal of the nucleosome hypothesis, according to which the elementary particle of chromatin (i.e., "the bead" in the electron micrograph) consisted of 180 bp of DNA combined with eight molecules of core histone and one molecule of linker histone.

D. The Structure of the Nucleosome

Data described in the preceding section set the stage for an experimental assault on the atomic structure of the nucleosome. Two lines of evidence obtained in the mid-1970s indicated that the histone proteins lie on the inside of the nucleosome, while the DNA is somehow exposed on its outside surface. M. Noll used the nuclease DNAse I to demonstrate that under appropriate experimental conditions, all 146 base pairs of DNA in a single nucleosomal particle are cleaved by this nuclease—one cleavage was seen to occur every 10–11 base pairs. This suggested that the nucleic acid is exposed to solution, rather than is shielded by the histone proteins. More definitive evidence to that effect came from biophysical studies in the labs of B. Richards and C. Crane-Robinson, who used neutron scattering to demonstrate that DNA does, indeed, lie on the outside of the core histone particle.

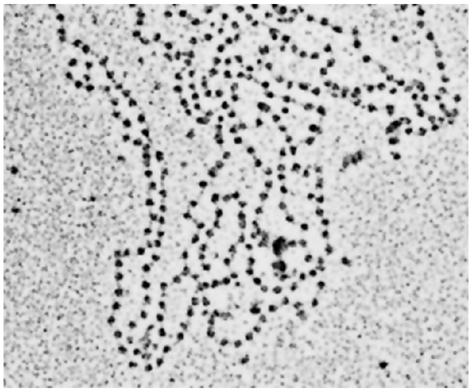

FIGURE 5 "Beads-on-a-string"—insect chromatin visualized under the electron microscope (EM kindly provided by U. Scheer).

In the 20 years that followed, the nucleosome was the subject of intense investigations. X-ray crystallographic analysis from T. Richmond and A. Klug, and subsequently G. Arents and E. Moudrianakis, illuminated the spatial arrangement of its constituents, as did protein-DNA cross-linking studies in the lab of A. Mirzabekov, while the details of the structural distortion that DNA undergoes in the nucleosome were provided by J. Hayes and A. Wolffe. The description that follows is based on data from all of these studies, as well as and the highest-resolution (2.8Å) X-ray crystal structure currently available (provided by T. Richmond's research group in 1997).

1. DNA Structure in the Nucleosome

Compaction of DNA into the nucleosome involves the winding of 146 base pairs of DNA into ca 1.7 left-handed turns around the histones (Fig. 6). Such a representation is very useful to help visualize what a nucleosome looks like but, unfortunately, presents the erroneous view that DNA is complacently wound onto the histones with little or no structural stress. The reality is quite contrary to what one could divine from this drawing: in assembling into the nucleosome, DNA is very severely distorted from its conventional and familiar B-form. First, to twist around the histones, the DNA backbone has to be very severely bent—the turns that it makes likely approach the limit of thermodynamic feasibility. In addition, topological requirements of winding a right-handed double helix into a left-handed superhelix necessitate that the DNA be partially unwound from its conventional 10.5 base pairs per helix turn.

The distortion of the DNA in the nucleosome has several important functional consequences. Because DNA needs to bend as it winds into the nucleosome, particular DNA sequences that bend more easily offer a thermodynamic advantage in this process; this means that the precise way in which a given DNA sequence associates with the histones—i.e., the way in which specific sequences in the DNA are rotated toward, or away from the core histones—

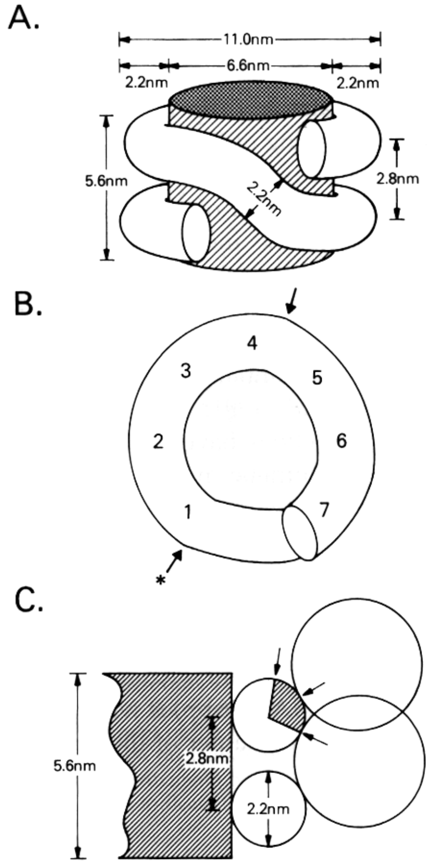

FIGURE 6 (A) A schematic of the DNA path around the nucleosome. (B) DNA distortion in one turn of the superhelix. (C) Steric hindrance in access to nucleosomal DNA.

is going to be at least in part sequence dependent. Such "rotational" positioning of the DNA relative to the core histone octamer can have profound effects on the regulatory behavior of the DNA in the nucleus, as we shall see.

Furthermore, the severe structural distortion of the DNA in the nucleosome is something of an Achilles heel, since it implies that this entire structure is amenable to disruption. We do not wish to create the impression that the nucleosome is intrinsically unstable—quite the contrary, an intact histone octamer can remain complexed with DNA under conditions of physiological pH and low ionic strength for very extended periods of time. Should the arrangement of the core histones within the nucleosome, or histone–DNA interactions themselves be altered, however, DNA will attempt to release topological and structural stress by recovering some B-form normalcy. This feature of the nucleosome is efficiently exploited in transcriptional control.

2. The Histone Octamer

The core histone octamer has a remarkably lucid structure; the key to appreciating this point—possibly hidden by the maze of barrels and ribbons in diagrams—is to realize that the globular domain of all core histones assumes an identical secondary and tertiary structure (Fig. 7). In this "histone fold" motif, the central, extended α helix is flanked on each end by a short turn (loop) of the polypeptide, and then—again, on each end—a shorter α helix (i.e., somewhat like the letter "Z"). The shorter helices reside on top of the same side of the longer helix—at its opposite ends, of course—and both lie at an approximately right angle to it. An identity of folding into this motif allows two histone molecules to heterodimerize via a "handshake": the central α helices of histones H3 and H4 fit together diagonally to form an "X" (extending the "handshake" analogy, this is equivalent to the touching of palms), and the shorter helices on each side of the central helix project toward their counterparts on the other histone molecule (analogous to fingers of one hand wrapping around the other hand). Histones H2A and H2B interact via an identical handshake.

In terms of binding DNA, the most immediate result of two histones coming together in a structure like this one is that the interhelical turns (loops) of each histone molecule become juxtaposed on each end ("top" and "bottom") of the "X" that is formed. These loops build a ramp onto which DNA is wound (each histone dimer associates with ca. 2.5 turns of DNA, i.e., ca. 27–28 base pairs); the shorter helices on one end of each histone project toward the "front end" of the "X" and also contact the DNA.

This histone heterodimer, then, is the elementary subunit of the nucleosome—the entire entity is built by fitting four such heterodimers together, in the following way: two H3/H4 heterodimers come together to form the H3/H4 tetramer (i.e., an "X_X" is assembled). Within the tetramer, the short α helix at the end of one histone H3 molecule contacts such a helix on the other histone H3. Thus, the histones are joined end-to-end: H4′–H3′—H3–H4. The resulting entity, in fact, does occur *in vivo* during postreplicative chromatin assembly, when it associates with approximately 120 bp of DNA. Its functional properties, however, differ quite significantly from that of the histone octamer bound to 146 base puirs; as shown in the lab of J. Hansen, DNA bound to an H3/H4 tetramer is much more accessible to binding by nonhistone regulators than DNA in a conventional nucleosome, and this has immediate functional consequences in terms of the effects of DNA replication on genome behavior (see following).

Once the H3/H4 tetramer is formed, it is joined by two dimers of H2A/H2B, primarily via hydrophobic and hydrogen bond contacts between histone H4 with histone H2B. The resulting daisychain of histones can be represented in linear form to illuminate the pattern of histone–histone interactions (contacts made within each dimer are

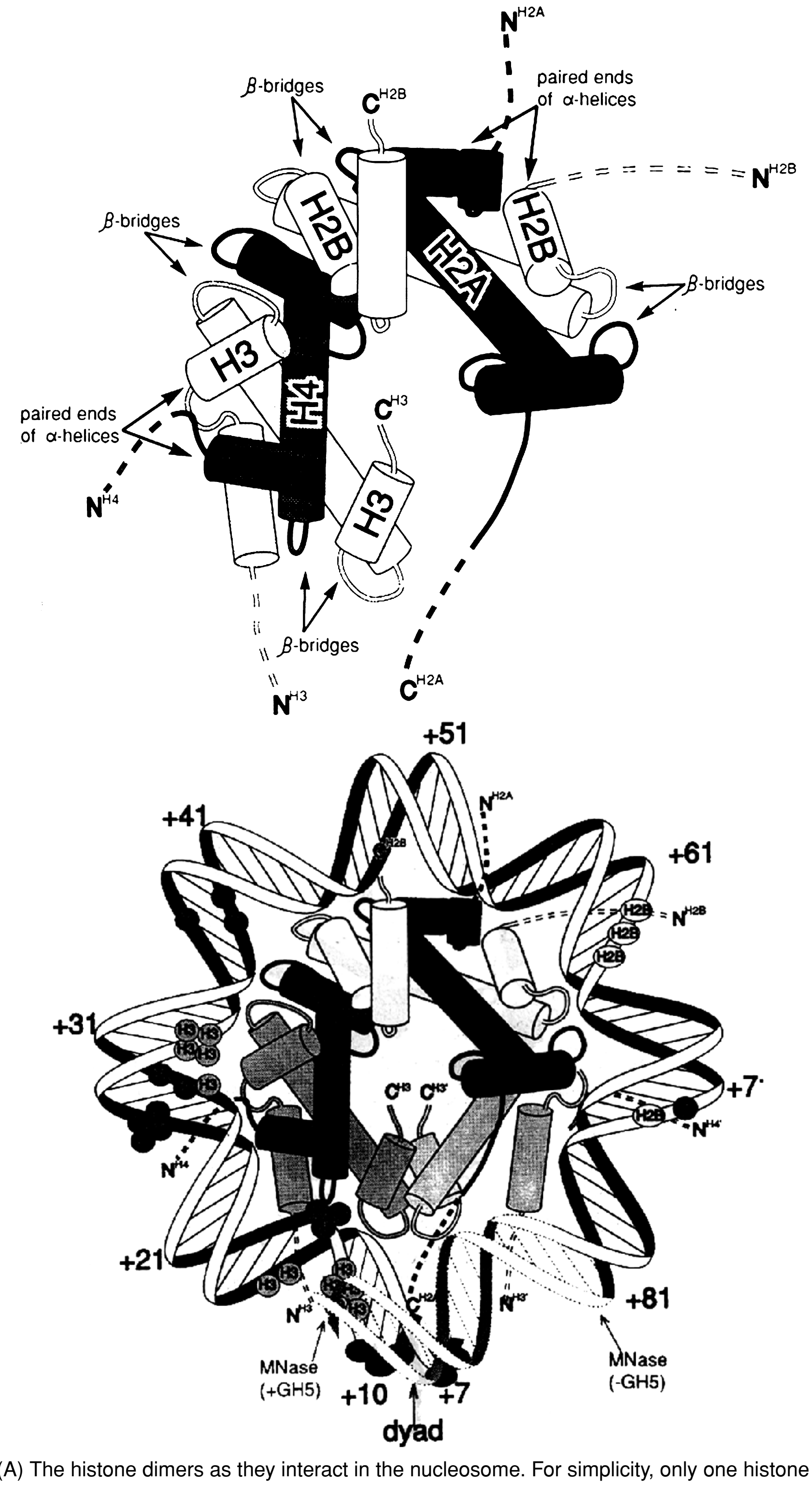

FIGURE 7 (A) The histone dimers as they interact in the nucleosome. For simplicity, only one histone set is shown. (B) The nucleosome and locations of histone–DNA interactions.

represented by a shorter dash, and between dimers—by a longer one):

H2A–H2B—H4–H3—H3–H4—H2B–H2A

In vivo, of course, the histones are not arranged in linear fashion, but rather bundle together into a rather compact particle: the H3/H4 tetramer forms an inverted V (i.e., a capital lambda, Λ), and the H2A/H2B heterodimers attach to either side of the cleft at the bottom of this "Λ." Viewed from the side, the resulting entity is shaped like a wedge (the octamer tapers toward the side of the H3/H4 tetramer). Yet another unfortunate consequence of textbook schematics is that the octamer is commonly represented as a hockey puck—i.e., somewhat of a monolithic entity. The octamer is not a monolith—in fact, it is very clear that the heterodimerization interface between histone H4 and histone H2B is a somewhat delicate one, and is amenable to disruption. Thus, a better analogy for the octamer would be that of a Rubik's cube—a three-dimensional jigsaw puzzle—with individual components fitting together in a dynamic, pliable arrangement.

Sixteen loop motifs decorate the sides of the octamer side like rungs on a ladder, and these offer an interaction interface to the DNA. The majority of contacts between the histones and the DNA occur via hydrogen bonds and salt links to the phosphates in the DNA backbone, although two additional important sets of interactions occur when the side chain of arginine in the histone penetrates the minor groove of the DNA, and when, remarkably, nonpolar contacts are made with the deoxyribose.

From a regulatory perspective, as shown in Fig. 6, a dramatic consequence of DNA winding onto the octamer is a change in its accessibility to solution (and, by extension, to nonhistone protein regulators). The obstacles are many in nature: the octamer itself obstructs the face of the DNA that is turned toward it, and—because of the proximity of the DNA gyres in the superhelix, one turn of the DNA impedes access to the adjacent DNA turn. A great deal of experimental attention has been directed, therefore, toward understanding how particular sequences bind to the histone octamer, and how specific regulators then bind to nucleosomes containing their target sites.

One important experimental variable in these assays cannot, unfortunately, be seen in the X-ray crystal structure: the NH_2-terminal core histone tails account for ca. 25% of the histone mass, but are mostly unstructured in the crystals. It is known that the tails of histones H2B and H3 do not stick out to the side of the nucleosome, but rather emerge into solution between the gyres of the DNA superhelix. The whereabouts of the histone H4 tail is unclear (with the exception of a small segment that makes an *internucleosomal contact* in the crystal!).

Attention to the tails' structure is justified by their prominence in regulatory phenomena that occur on chromatin (Section IV.A). While normal nucleosome assembly can occur on tailless (proteolyzed) histones, normal transcriptional control can not—and the regulatory reach of the tails can be illustrated by their sheer stereochemical reach (Fig. 8). Because the tails contain a combined 44 lysine residues, at physiological pH they very likely coat the

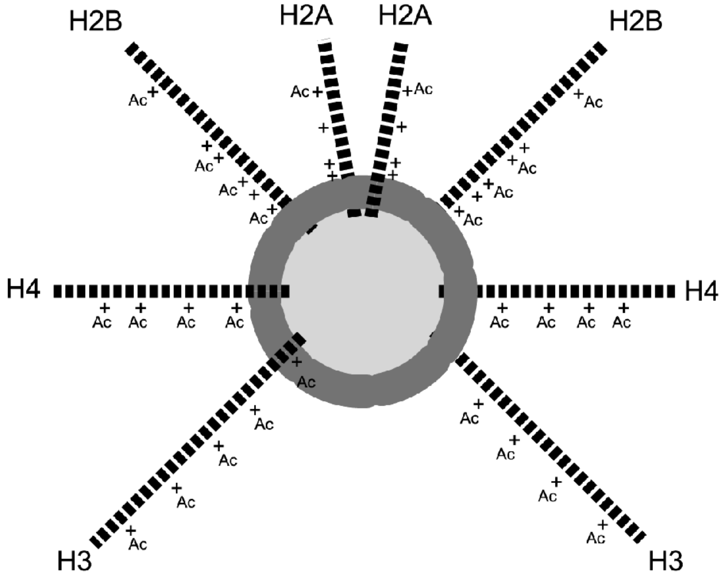

FIGURE 8 The histone tails shown schematically to their full predicted length in relation to the nucleosome core particle (drawn to scale).

underlying DNA like a maze of positively charged tentacles. Thus, the alterations in the charge of the tails effected by postranslational covalent modifications are likely to have impact on chromatin.

E. "... Not by Beads Alone": Higher-Order Chromatin Structure

It is clear that the nucleosome as a tool is not sufficient to compact the entirety of the genome into the nucleus; quite a feat of condensing is required to convert the "beads-on-a-string" fiber into the dramatic metaphase chromosomes so familiar to many (if the entirety of the genome were assembled into a fully extended nucleosomal array, its length would be ca. 15 cm).

One important functional component of further folding by the nucleosomal fiber is the linker histone (H1). As its name implies, its binding site is with the DNA stretch between two adjacent octamer particles; the precise manner in which the linker histone binds to that DNA has been the subject of much investigation and controversy (Fig. 9). Whatever the precise mode of its association with DNA, one important consequence of this association is charge neutralization over the linker DNA stretch—that is, the shielding of the negatively charged phosphate backbone from solution. Experiments in the laboratory of M. Gorovsky have shown that the ciliate *Tetrahymena* can survive without linker histone, but that the size of its nucleus increases twofold; furthermore, mutations that increase negative charge on histone H1 (mimicking its phosphorylation) promote its loss from chromatin. In addition, data on the folding of chromatin *in vitro* in the presence and absence of linker histone are also fully consistent with the model that the presence of histone H1 is required for proper folding of the "bead-on-a-string" fiber.

What exactly this fiber folds into is—remarkably—not known. A great variety of EM and other approaches—dutifully represented in textbook schematics—have suggested that the next level of chromatin compaction is an entity termed "the 30-nm fiber." In 1979, F. Thoma, A. Koller, and A. Klug proposed a model for this entity according to which the nucleosomal fiber winds onto itself to form a solenoidal structure, with ca. six nucleosomes per turn of the superhelix. Remarkably, whether such an entity forms *in vivo*, and what the precise arrangement is of the nucleosome within this structure, remains an open issue, and other models for this fiber have been proposed by C. Woodcock and several other scientists.

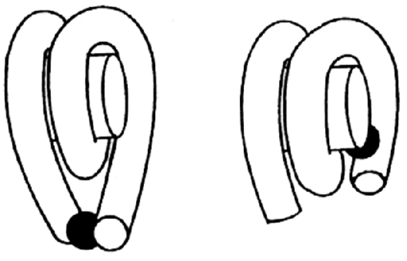

FIGURE 9 Two proposed models for the location of the linker histone (black sphere) relative to the nucleosome core particle.

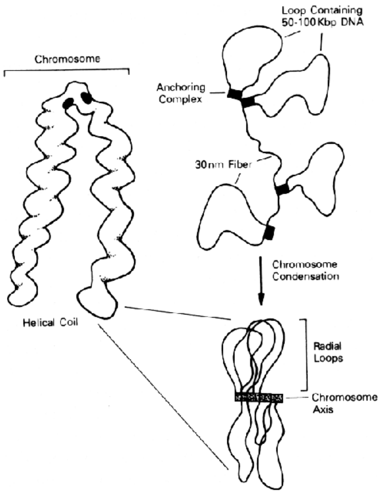

FIGURE 10 Model for higher-order chromatin folding.

Beyond the mysterious 30-nm fiber lies an undiscovered country of dramatic proportions—we currently lack the technical tools to examine higher-order chromatin folding. It has been proposed (Fig. 10) that large domains of chromatin emerge from a central scaffold in the form of loops, and work from the labs of U. Laemmli and J. Sedat presented evidence in support of this model. Details remain very elusive, however. It is important to realize, however, that even in the absence of information about higher-order chromatin structure, our current understanding of the nucleosome and the nucleosomal fiber offers an ample stage for the unfolding of very complex gene regulatory phenomena concomitant with chromatin structure transitions. These are reviewed in the next section.

IV. THE DISRUPTION AND MODIFICATION OF CHROMATIN STRUCTURE AS A TOOL TO CONTROL THE GENOME

It is somewhat ironic that the discovery of the nucleosome, and the determination by M. Noll in 1974 of its ubiquity in the genome, were partly responsible for the transient elimination of histones from the stage on which transcriptional control was thought to unfold. In retrospect, it is easy to see why: because all nucleosomes are the same (in a certain sense, they are), it was very hard to imagine,

how anything pertinent to *gene-specific* regulation could occur on such a homogeneous, monotonously reiterative substrate as the "beads-on-a-string" fiber. Thus, all transcriptional and other regulatory phenomena were thought to occur on nucleosome-free stretches of "naked" DNA, although it remained unclear, whether such stretches can be found *in vivo.*

This notion was also supported by experimental observations on nonhistone factor interactions with chromatin templates *in vitro*. Such analysis indicated that many proteins—for example, the TATA-box binding protein, TBP—cannot bind to their DNA sites when they are assembled into nucleosomes. From these, and other observations, emerged a model according to which inactive regions of the genome were packaged into nucleosomes; chromatin, thus, was viewed as a general (nonspecific), repressive entity. The elimination of histones from active regions of the genome was then thought to set the stage for binding by nonhistone factors and transcriptional activation via the recruitment of RNA polymerase.

Over the past years, however, several lines of evidence emerged that suggested this model was an oversimplification. It became apparent that chromatin structure is not monotonous or homogeneously isomorphous; instead, extensive localized alterations in its nucleoprotein fiber were observed concomitant with changes in genomic activity at specific loci. In an important complementary development, a host of macromolecular complexes were discovered that populate the nucleus and effect these alterations. In addition, specific gene loci were shown to have proper transcriptional control *depend* on their assembly into chromatin. Finally, whole-genome expression profiling experiments described above provided strong evidence that chromatin is not a generalized repressor, and that many genes require chromatin for proper regulation.

A. Chromatin Structure Alterations That Occur *in vivo*

While structural studies of the nucleosome provided very important information for scholars of transcription, a major issue for understanding how the genome behaves in chromatin form was to learn what—if anything—happens to chromatin during gene activation and repression *in vivo*. The cytological studies described earlier in this article (Section II) indicated that localized alterations do occur, but the resolution limitations inherent in such methods prevented an interpretation of the data in molecular terms: for example, puffs that form on polytene chromosomes are very conspicuous, but what exactly happens to the chromatin fiber during puffing remained unclear (and, incidentally, still does).

1. Remodeling

As was the case with the initial identification of subunit composition of chromatin, the first clues to chromatin remodeling phenomena concomitant with alterations in gene activity came from the use of nucleases. In the late 1970s, C. Wu and S. Elgin, and, independently, S. Nedospasov and G. Georgiev used nucleases such as DNAse I and micrococcal nuclease (MNase) to probe chromatin *in vivo*. The notion behind this approach was the use of *low* quantities of enzyme—the prediction being that mature chromatin, i.e., DNA that is tightly complexed with histones, would not be accessible to cleavage by nuclease, and thus *only DNA stretches that were not assembled into conventional chromatin would be "visible" to the nuclease*. These investigators used an elegant technical trick—"indirect end-labeling"—to reveal such DNA stretches. Wu and Nedospasov made the same observation—that regulatory DNA (for example, promoters of active genes, or of genes that are poised for upregulation) is preferentially accessible to such nucleases and forms a "nuclease hypersensitive site" *in vivo*. At the time, the structural basis of this entity was unclear; remarkably, it continues to be not entirely certain to this day, although we have a much deeper understanding of the molecular machines responsible for effecting this chromatin structure alteration and of the properties these machines exhibit *in vitro*.

An appreciation for the role of such hypersensitive sites to gene regulation grew in the early 1980s, when it was discovered that they are a relatively ubiquitous feature of promoters and enhancers in the eukaryotic genome. A seminal observation was made in 1984 using a model system that has provided many insights into the role of chromatin in gene control: the regulation of transcription of the mouse mammary tumor virus (MMTV) genome by the glucocorticoid receptor (GR) and its ligands, the glucocorticoids (cortisol, corticosterone, and aldosterone). GR is a small-molecule-regulated transcription factor: inactive in the absence of hormone, it translocates to the nucleus in its presence, binds to target genes, and activates transcription. In 1974, it was discovered that transcription driven by MMTV—the etiologic agent of breast cancer in mice—is rapidly upregulated by treatment with a synthetic GR ligand, dexamethasone. Ten years later, K. Zaret and K. Yamamoto discovered that hormonal treatment induces a strong DNAse I hypersensitive site in the MMTV promoter (called the "long terminal repeat," or LTR), and that this hypersensitive site vanished rapidly upon removal of hormone. This established a correlation between the extent of such remodeling and the level of transcription at this locus. An example of a DNAse I hypersensitive site induced by a nuclear hormone receptor is shown in Fig. 11.

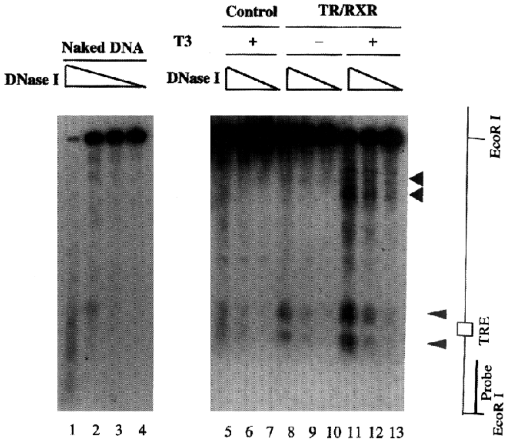

FIGURE 11 Binding of the thyroid hormone receptor (TR) to chromatin generates a DNAse I hypersensitive site (arrowheads). "TRE" represents the TR binding site.

It is important to note that the disruption of chromatin structure that is described as a "hypersensitive" site is a highly localized event and only affects ca. 50–500 base pairs of DNA (i.e., only several nucleosomes' worth)—thus, while the insect relative of GR, the ecdysone receptor, functions via highly similar mechanisms, it should not be inferred that polytene chromosome puffing observed some 25 years prior by Clever and Karlson in *Chironomus* after ecdysone injection (see Section II) is due to such remodeling—puffing is a large-scale event which affects stretches of the genome that are many thousands of bp long. A molecular correlate to "puffing" was discovered in the lab of C. Crane-Robinson in 1994 in studies of the globin gene cluster. A marvel of gene regulation, the globin genes in higher vertebrates have been investigated in great detail, in part because they represent one of the best characterized instances of a "chromosomal domain"—a large continuous stretch of the genome in which several genes reside whose expression is coordinately regulated (for instance, in mammals, a progression of gene activation occurs, with "fetal" globin genes active during embryonic development, and a switch to "adult" type globin genes after birth). In the chicken genome, the β-globin locus encompasses ca. 33,000 base pairs; by using nucleases, Crane-Robinson's research group showed that in erythrocytes (i.e., the cell type that transcribes globin genes) this entire portion of the chromosome was more sensitive to nucleases than a transcriptionally inert one.

We emphasize that the distinction between nuclease *sensitivity* and *hypesensitivity* is not merely a semantic one. During the activation of gene expression *in vivo*, it is quite likely that as a first step, a large stretch of chromatin undergoes some structural transition to a more accessible conformation (it is possible that such a transition is mechanistically similar to that between heterochromatin and euchromatin). Within such a large domain of less compacted chromatin, targeted chromatin remodeling over short DNA segments then induces *hyper*sensitivity, i.e., a more dramatic disruption of histone–DNA contacts.

2. Modification

An additional important observation on the β-globin locus related to a different type of chromatin structure alteration observed *in vivo*: a change in the covalent modification status of histone tails within a domain of transcriptionally active chromatin. To better illuminate the significance of this finding, we must briefly review the biochemistry of such modifications. As mentioned earlier, in 1964 it was discovered that particular lysine residues in the NH_2-terminal tails of the core histones are reversibly covalently modified by acetylation:

$$\cdots\text{—}NH_3^+ + \text{Ac-CoA} \rightarrow \cdots\text{—}NH_2\text{—}CO\text{—}CH_3.$$

This observation instantly suggested an electrostatic mechanism for gene regulation: the histone tails carry a wealth of positive charge on their basic amino acids and are thus expected to bind tightly to the negatively charged phosphates in the DNA backbone. Acetylation eliminates the charge on the target amino acid side chain, and thus a hyperacetylated histone tail would be expected to bind less tightly to the DNA, and make it more visible to the intranuclear world. Thus, there must be a correlation between chromatin hyperacetylation and transcriptional activation (and, conversely, between chromatin deacetylation and transcriptional repression).

This model makes a lot of sense from first principle *a priori* considerations, but is it supported by experimental evidence? The correlation between levels of acetylation and transcription clearly exists *in vivo*. A very powerful tool in making this experimental determination has been antibodies that selectively recognize hyperacetylated or deacetylated histone tails. These have been used with great success in two experimental strategies. The first, fluorescent *in situ* hybridization (FISH), is cytological: it allows one to examine histone acetylation status over entire chromosomes. In the most general sense, it involves the preparation of nuclei from cells (or tissue) under biochemically mild conditions, the immobilization of their chromatin content on a suitable support, and the probing of the resulting karyotype with an antibody of choice, followed by probing with a "secondary" antibody that reveals where the original antibody bound. The secondary antibody is chemically coupled to a fluorescent dye, and this allows visualization under a microscope of entire chromosomes with brightly fluorescent segments that have bound to the antibody.

A model system best suited for such analysis is one where an entire chromosome has an altered expression state; the best characterized example comes from studies of dosage compensation: this evolutionarily conserved device for coping with a different autosome to sex chromosome ratio between genders in metazoa changes expression levels of the X chromosome depending on gender. For example, in mammals, one of the two X chromosomes in females is inactivated (thus, identical expression levels are achieved for X-linked loci between males and females). In insects, on the other hand, the single X in males is transcriptionally upregulated twofold, thus adjusting its expression level to that of two X chromosomes in females. FISH analysis by B. Turner and colleagues demonstrated that the inactive X in mammalian females is hypoacetylated; this provides an important correlate to Ohno's and Lyon's observations (v.s.) from the 1960s that this chromosome is condensed into heterochromatin and transcriptionally silenced. On the other hand, FISH analysis in *Drosophila* revealed that the transcriptionally "hyperactive" X chromosome in males is hyperacetylated relative to X chromosomes in females.

The second experimental approach that revealed a correlation between states of acetylation and levels of transcriptional activity is chromatin immunoprecipitation (ChIP). This method was developed by M. Solomon and A. Varshavsky, and further by D. Allis and M. Gorovsky. This technique allows one to detemine if a protein of interest interacts *in vivo* with a DNA stretch of interest; the reagents required for such analysis are an antibody against a particular protein and knowledge of the primary DNA sequence of the locus of interest. This ingenious and powerful method begins by taking an *in vivo* snapshot of protein–DNA and protein–protein interactions in the nucleus via a brief incubation of living cells or tissue in formaldehyde; this small molecule rapidly penetrates into the cell and introduces covalent crosslinks between proteins and DNA that they are bound to, as well as between proteins that are in sufficient physical proximity (i.e., are in a complex). Chromatin is then isolated from cells and sheared (by acoustical means, i.e., sonication) into small—ca. 500 base pairs—fragments. An immunoprecipitation is then performed to isolate from this complex mixture the protein of interest (and whatever happens to be covalently attached to it): the antibody is immobilized on a suspension of agarose beads, and the beads are then mixed extensively with the sonicated chromatin to allow the antibody to bind its antigen. The beads are then isolated by centrifugation, which redistributes the target protein from solution into the pellet (together with the beads). The cross links between the protein and DNA are then eliminated, and the DNA isolated. Finally, the presence of a given DNA sequence in this isolate is assayed by PCR or by Southern blotting.

ChIP has been applied extensively to analyze the histone tail acetylation status over particular stretches of various genomes (from budding yeast to humans). The general conclusion from these experiments is that, indeed, transcriptional repression is accompanied by localized histone deacetylation, while transcriptional activation occurs within loci that are associated with hyperacetylated histones: for example, theDNAse I sensitive domain of the active chicken β-globin locus that was described earlier was found to be hyperacetylated by ChIP analysis (C. Crane-Robinson and colleagues), while transcriptionally silent stretches of yeast chromatin, such as the mating type loci, are deacetylated (J. Broach and colleagues).

It is quite striking that our current biochemical insight into the enzymatic reaction of histone tail acetylation, the causative agents of this modification (Section IV.C), and their involvement in transcriptional control *in vivo* is not paralleled by a similar understanding of the structural effects of histone tail hyperacetylation on chromatin, or of the mechanistic underpinnings of the general stimulatory effect that this modification has on the transcriptional machinery.

The structural puzzles come in part from the fact that the tails fail to appear in X-ray crystallographic analysis. It is clear that in the case of histones H3 and H2B, the tails emerge into solution by passing through the two adjacent DNA double helices that lie on the surface of the octamer, but their subsequent path—assuming a defined one exists, which is not at all clear—is unknown. In addition, a short segment of the histone H4 tail—seven amino acids—can be seen making contact with a histone H2A/H2B dimer in an adjacent nucleosomal particle. How—or whether—the tails engage the DNA of the nucleosome they belong to is unknown.

Faced with a void of structural understanding from crystallographic analysis, scientists turned to other biophysical methods to investigate what happens to the tails upon acetylation. In 1982, E. M. Bradbury and colleagues used NMR analysis of peptides corresponding to the histone H4 tail, and found that it bound only weakly to DNA, and that hyperacetylation abolished this binding. By thermal denaturation analysis these scientists derived a quantitative estimate of the effect of acetylation: the intact peptide was seen binding to DNA with an affinity of 50 pM, while acetylation reduced it to 10 μM! The magnitude of this effect was subsequently shown to depend on the number of lysine residues acetylated, and in the context of the nucleosome—rather than as isolated peptides—the histone tails continued to make some contacts with the DNA even when hyperacetylated. In studies using circular dichroism spectra (J. Parello and colleagues), DNA-bound stretches of both histone H4 and H3 tails were found to be highly structured and adopt an α-helical conformation,

while those of histones H2A and H2B were seen to assume a random coil conformation. The working model as to the local effects of acetylation on tail structure and interaction with DNA, then, suggests that it may disrupt the (currently unknown) secondary structure the tails assume, and lessen the extent to which the tails interact with DNA.

What is the relevance of these observations to transcriptional control? If, indeed, the tails become more loosely associated with DNA upon hyperacetylation, it is possible that the underlying DNA becomes more accessible to nonhistone regulators. *In vitro* experiments with purified chromatin components and particular transcriptional regulators (A. Wolffe, J. Workman, and their colleagues) have found that histone hyperacetylation potentiates binding to nucleosomal substrates by such proteins as TFIIIA, Gal4, and USF. Whether such potentiation of binding occurs *in vivo* is unknown.

As mentioned earlier, a segment of the histone H4 tail was seen making an *internucleosomal* contact in the crystal structure. These data, and other observations, lent support to the hypothesis that tail acetylation affects higher-order folding of chromatin. Strong evidence to this effect was obtained by J. Hansen and his colleagues, who used analytical ultracentrifugation to demonstrate that the hydrodynamic properties of chromatin fibers can be modulated *in vitro*. An alteration in the ionic strength of the solution dramatically altered the shape of the nucleosomal fiber: as the concentration of Mg^{2+} increased, the fiber became much more compact. Importantly, the deacetylation of this fiber was then shown to have an identical effect: thus, hyperacetylated chromatin adopts an extended conformation, and deacetylation promotes folding. It is possible, therefore, that the increased accessibility to nucleases seen in transcriptionally active, hyperacetylated chromosomal domains *in vivo* is causally linked to a change in the extent of chromatin compaction that can be observed *in vitro*. Additional support for such a connection comes from the work of A. Belmont and coworkers, who used *in vivo* imaging techniques to demonstrate that transcriptional activation is accompanied by the hyperacetylation and unfolding of a chromosomal domain spanning several hundred thousand base pairs. Most remarkably, this process occurred even in the absence of transcription—this important control experiment indicated that the unfolding is not a consequence of the passage by the RNA polymerase II complex (an imposing entity) through the chromosome, and is an independently regulated phenomenon.

Whatever the structural consequences of acetylation on chromatin, some of the strongest evidence in favor of its direct role in controlling the genome comes from data illuminating the abundance inside the nucleus of enzymes effecting histone tail modification, and from experiments that show these enzymes to be directly involved in transcriptional regulation *in vivo*. These are reviewed in the next section.

B. Chromatin Disruption and Modification: The Enzymatic Machinery

In interesting testimony to the powerful influence of methodology over the course scientific inquiry, investigations of the molecular agents that effect chromatin structure alterations *in vivo* have followed a remarkably uniform scheme, somewhat reminiscent of a *Bildungsroman*. It begins with a genetic screen in budding yeast for strains that exhibit particular phenotypes related to gene control (e.g., the ability to activate a particular metabolical pathway), the molecular cloning of the underlying loci, and their subsequent putative identification as transcriptional regulators. Next, analysis in metazoa reveals the existence of homologs, and an enzymatic assay is developed that demonstrates that these proteins have the capacity to alter chromatin structure. Mutational analysis reveals that there is a correlation between the enzymatic activity of the protein and its ability to act in transcriptional control. Biochemical analysis of whole-cell extracts then finds these proteins within large, multisubunit complexes, only a few polypeptides in which have enzymatic activity, while the rest are of uncertain function. Finally, *in vitro* experiments are done that show these complexes can be targeted by particular transcriptional activators or repressors, and that mutations in these transcription factors that alter the ability to interact with chromatin modifying and remodelling complexes impair their properties as regulators.

1. Disrupt and Conquer: ATP-Dependent Chromatin Remodeling Engines

In the 1980s, genetic studies in laboratories of M. Carlson and F. Winston identified a number of mutant yeast strains that failed to metabolize sucrose; following convention, the many loci revealed in the screen were called *SNF* (sucrose nonfermenter) and given a number (i.e., *SNF2, SNF5,* etc.). At approximately the same time, the lab of K. Nasmyth discovered a number of loci in the yeast genome that when mutated, incapacitated the switching of mating type; these were dubbed *SWI* (for "switch") and also numbered (*SWI5, SWI2*, etc.). The capacity to metabolize sucrose is dependent on the activation of specific genes that belong to the cognate metabolic pathway (in particular, an invertase); the switching of mating type requires the upregulation of a gene that encodes an endonuclease required for initiating a DNA recombination reaction. Thus, it appeared that a common feature of the two otherwise quite unrelated phenotypes in the mutant strains (failure to metabolize sugar or the incapacity to mate) was a

deficiency in gene activation; most remarkably, the same locus was commonly found mutated in both screens. The cognate gene—*SWI2/SNF2*—was subsequently found in work from I. Herskowitz's laboratory to be required for transcriptional upregulation of a number of budding yeast genes.

In a landmark set of experiments, the lab of F. Winston in 1992 demonstrated that the action of *SWI2/SNF2* in gene activation is mediated via alleviating the repressive effect on gene expression of chromatin structure. These studies used the classical genetic notion of epistasis—i.e., the capacity of mutations in one locus to mask a mutation in a different locus—to uncover genetic interactions between *SWI2/SNF2* and the genes for histones: it was shown that the inability of certain genes in the yeast genome to be activated when *SWI2/SNF2* is mutant can be "healed" by making mutations in histone genes (for example, by lowering the histone content of the nucleus). Most strikingly, subsequent analysis by I. Herskowitz, C. Peterson, and M. Osley demonstrated that even less drastic measures—for example, point mutations in histone H4 that impair the capacity of H2A–H2B to bind the (H3/H4) tetramer—will also make *SWI2/SNF2* unnecessary. Thus, those genes in yeast that require *SWI2/SNF2* to become transcriptionally active *lose that requirement* when chromatin structure is destabilized by making mutations in histone genes.

The most immediate prediction of these remarkable experiments—consider the extraordinary fact that the *entirety of chromatin within the yeast nucleus can be altered by genetic means*—is that the product of the *SWI2/SNF2* gene somehow alters chromatin structure over target gene promoters. A great number of studies have yielded data fully compatible with that notion (the laboratories are too many to list, but, in addition to those already mentioned include those of G. Crabtree, W. Hörz, R. Kingston, C. Peterson, J. Workman).

Several important facts emerged regarding *SWI2/SNF2*. Biochemically, it was found to be an ATPase (i.e., it hydrolyses ribo-ATP to release ADP and phosphate)—this was an important observation, because it illuminated a possible requirement for energy in chromatin remodeling. *In vivo*, it was discovered as one of the core components of a multisubunit complex designated as SWI/SNF (pronounced "switch-sniff"). *In vitro* assays with nucleosomal templates and transcriptional regulators demonstrated that SWI/SNF can remodel histone–DNA contacts such that subsequent access by these regulators to the remodeled nucleosomal template is increased. For example, a TATA box buried within a nucleosome became more accessible to TBP after transcriptional activator-dependent action on that nucleosome. The exact structural nature of the "remodeled" nucleosome is unclear; it is known that the histone–DNA contacts are loosened sufficiently to allow access and cleavage by DNAse I, but it is also clear that the histones remain in some contact with the DNA. Whatever the nature of the remodeled entity, energy derived from the ATP hydrolysis is required for its generation.

What relevance does such action of SWI/SNF *in vitro* have to transcriptional control *in vivo*? The best explanation we can offer comes from studies in mammalian systems. Thus, work by G. Hager and colleagues on transcriptional control of the MMTV promoter (Section IV.A.1) focused on the function of the DNAse I hypersensitive site that is induced by the liganded glucocorticoid receptor concomitant with transcriptional activation. An important clue came from a comparative analysis of MMTV transcription on copies of the viral genome that have been integrated into the chromosome (and thus assume native chromatin organization) vs. such DNA that has transiently introduced into the cell by a technique called transfection (the DNA remains extrachromosomal and does not assemble physiological chromatin; it is lost from the cell after only a few rounds of cell division). It was known that full-scale activated transcription on the MMTV promoter required an activator called NF1; remarkably, NF1 action was only dependent on GR and its ligand when the target promoter was chromatinized: on transiently transfected DNA, NF1 was able to bind to DNA without any abetting action from the receptor.

An explanation for this interesting synergy was provided in a famous study by T. Archer and G. Hager in 1992: they proposed that the MMTV promoter adopts a *nonrandom* chromatin organization *in vivo*, such that the binding site of NF1 is occluded by a nucleosome. In contrast to NF1, GR can directly bind to chromatin over the MMTV promoter, and then somehow remodels histone–DNA contacts adjacent to its binding sites. This remodeling (manifested as a DNAse I hypersensitive site) facilitates NF1 access and potentiates transcriptional activation (Fig. 12). This two-step ("bimodal") mechanism for GR action led to several predictions: the receptor had to be shown as competent for binding to nucleosomes, and also for the recruitment of a chromatin remodeling engine. In fact, GR and several other members of the nuclear hormone receptor superfamily can bind to nucleosomes *in vitro*—this is quite an achievement, considering the extensive steric hindrance exerted by the nucleosome (an example of such binding to nucleosomes by the thyroid hormone receptor is shown in Fig. 13). In addition, T. Archer and colleagues demonstrated that a human homolog of the budding yeast SWI/SNF complex is required for transcriptional activation of MMTV by GR. A central conclusion of this analysis is that *bona fide* transcriptional control of this promoter cannot be recapitulated on naked DNA, and that the infrastructure of chromatin is integrated into the transcriptional regulatory pathways affecting MMTV.

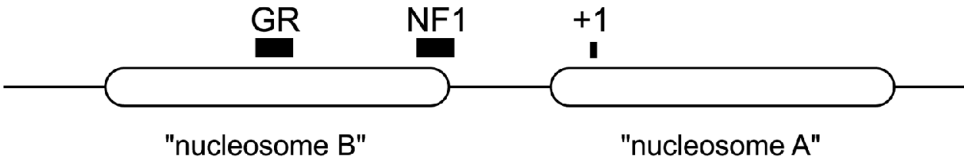

FIGURE 12 A schematic of the low-resolution map of the MMTV promoter (LTR); the position of the transcription start site ("+1"), the GR binding site, the NF1 binding site, and the histone octamers are shown.

Whether SWI/SNF is directly responsible for the chromatin remodeling over this, or any other mammalian promoter *in vivo* remains an open issue. Work from O. Delattre and colleagues showed that truncating mutations in a core subunit of the human SWI/SNF complex cause a devastating disorder—"malignant rhabdoid tumors" (MRT); these develop on the kidneys and in the central nervous system in infants (onset occurs before 2 years of age) and are highly aggressive. This indicates that the maintenance of normal cell proliferation and differentiation status in humans require the function of the SWI/SNF complex. It is not clear, however, if the facilitation of access via the remodeling of histone–DNA contacts accounts for the entirety of action by SWI/SNF during activation.

For specific promoters in budding yeast, it has been shown that ablation of the SWI/SNF complex causes these promoters to become inactivated due to a failure to remodel chromatin: no DNAse I hypersensitive site is visualized over these promoters in the absence of SWI/SNF. One mysterious issue concerning the function of SWI/SNF is its role in transcriptional repression; whole-genome expression analysis in the laboratory of R. Young demonstrated that a number of budding yeast genes become transcriptionally activated in strains lacking *SWI2* (and a different set of genes become silenced). While the phenomenon is clear, its mechanistic underpinnings are not. Elegant genetic analysis in F. Winston's laboratory demonstrated that SWI/SNF acts directly on those promoters that are upregulated in its absence (i.e., it is a direct repressor of those genes, and not, for example, an activator of a repressor). It is possible that the repressive action of SWI/SNF has to do with its ability to effect nucleosome mobilization: the sliding of the histone octamer *in cis* relative to the DNA. Thus, perhaps, SWI/SNF actively repositions nucleosomes over particular gene promoters such that important regulatory DNA stretches are occluded.

Concluding our overview of ATP-dependent chromatin remodeling machines, we note that eukaryotic genomes are populated with relatives of the *SWI2/SNF2* ATPases (the best characterized such relative is the protein ISWI—pronounced "eye switch"). These also occur in a large, multisubunit complexes, but their role is unclear. Many of these complexes have the interesting ability to organize chromatin—i.e., introduce proper spacing into disordered

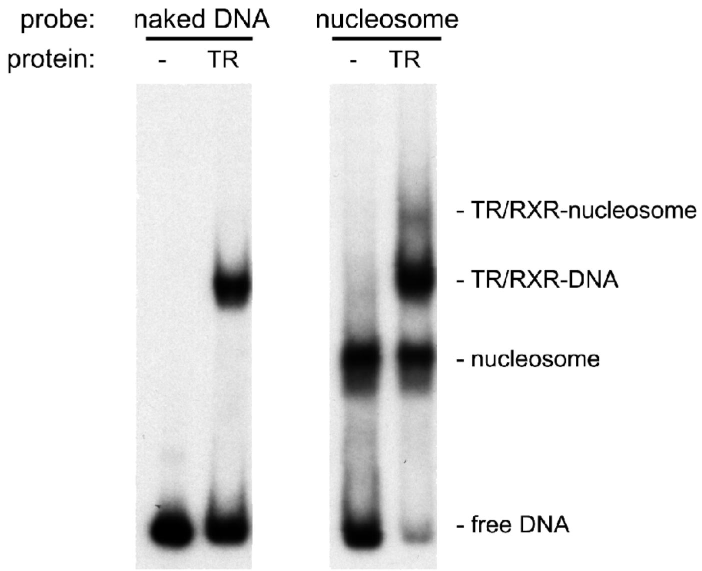

FIGURE 13 The thyroid hormone receptor can bind to naked DNA containing its response element (left half), and also to this DNA when it is assembled into a nucleosome (right half).

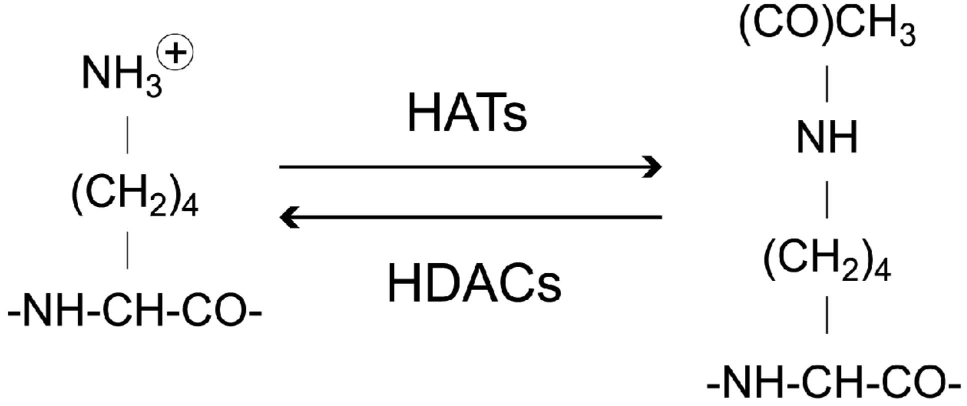

FIGURE 14 The chemical equation of ε-acetyllysine synthesis and degradation by HATs and HDACs.

nucleosomal arrays—but what these complexes actually do to the chromatin fiber *in vivo* remains to be investigated.

2. The Power of Amendments: HATs and HDACs

The extensive evidence connecting histone tail acetylation with transcription (Section IV.A.2) suffered from its inherently correlative nature—it was not clear, whether changes in acetylation status were causative to, or, instead, an effect of, changes in gene activity. This situation changed quite dramatically in 1996.

As shown in Fig. 14, the reaction of acetylation is catalyzed by histone acetyltransferases (HATs) and reversed by histone deacetylases (HDACs). In the cases of both classes of proteins, their enzymatic activity was used as a tool to purify them from cells. The laboratory of C. D. Allis developed a very unusual method for the detection of HAT activity: the in-gel assay. In this technique, a conventional polyacrylamide gel is prepared with a large quantity of histone protein polymerized directly into the gel matrix! A crude protein mixture is then resolved on the gel, and then an *in loco* reaction is performed by bathing the gel in ^{3}H-acetyl-CoA: the assumption is that the presumptive HAT will, after having migrated to a defined position in the gel, utilize the histone substrate all around it in the gel, and label it. The position of the HAT activity, therefore, can be visualized by fluorography. Allis's lab identified a major source of HAT activity in nuclei of the ciliated protozoan *Tetrahymena* that migrated at 55 kDA; its purification and primary sequence characterization yielded the striking observation of its high homology to a known yeast protein, Gcn5p.

To appreciate the significance of this discovery, it is helpful to briefly review what was then known about *GCN5* (in budding yeast, gene loci are designated by uppercase italics, i.e., *ABC1*, while the protein product of that locus is written as "Abc1p"). This gene was isolated in the mid-1980s in a screen for strains that would have deficiencies in regulating enzymes involved in amino acid biosynthesis; nothing was known about the mechanism of its action until the lab of L. Guarente demonstrated in 1992–1994 that it is involved in transcriptional control by such transcription factors as Gal4p and Gcn4p. The genetic approach that led to this discovery is very informative: the chimeric transcriptional activator Gal4-VP16 carries the DNA binding domain of the yeast transcription factor Gal4p, and the transcription activation domain of a protein from herpes simplex virus (HSV)—virally derived proteins and regulatory DNA stretches have evolved to evoke large-scale responses from the transcriptional machinery, and thus are very commonly used in biology. An unexpected property of Gal4-VP16 was its toxicity to the yeast cell; work from M. Ptashne's lab gave rise to the notion of "squelching"—competition by a given transcription factor for some limiting general component of the basal transcriptional machinery. Thus, it was thought that thte VP16 domain "squelches" important regulatory proteins away from other yeast promoters. Thus, reasoned Guarente and his colleagues, whatever protein—when overexpressed—suppresses Gal4-VP16 toxicity must be relevant to transcriptional control. Indeed, such a screen yielded the *GCN5* locus; it was suggested that Gcn5p, therefore, is a "transcriptional adaptor"—an entity that transduces the regulatory signal being sent by the DNA-bound regulator to the basal transcription machinery.

Thus, the discovery of Allis and coworkers shone in a new light: a protein found in ciliates enzymatically capable of hyperacetylating histones has a very close relative in budding yeast—a protein that is required for transcriptional activation and is thought to be an adaptor! This was the first instance of a direct connection between a specific HAT and transcriptional control. In subsequent years, Gcn5p in yeast and other organisms has been subjected to an extraordinarily comprehensive set of studies that firmly

established its HAT activity as essential to transcriptional activation. For example, work from Allis and coworkers showed that *GCN5* is required for the transcriptional upregulation of a number of budding yeast genes, and that point mutations in Gcn5p that abrogate its catalytic activity as a HAT abolish its capacity to abet transcriptional control. In addition, ChIP analysis demonstrated that histones over promoters of genes that are targets for binding by Gcn5p-recruiting activators are hyperacetylated concomitant with its action. Finally, the laboratory of S. Roth reported an experiment that is the scientific equivalent of a *coup de grace*: a yeast strain was engineered in which specific lysines in histone tails were mutated to a noncharged residue—thus, these cells had "genetically hyperacetylated histones." In such a strain, formerly *GCN5*-dependent genes lost their requirement for Gcn5p to become transcriptionally active!—thus, it was formally proven that the *in vivo* function of Gcn5p at target gene promoters is to hyperacetylate the histones.

While original analysis revealed a single major HAT in *Tetrahymena* extracts, subsequent work revealed the fact that eukaryotic genomes contain a large number of proteins possessing HAT activity. Importantly, an overwhelming majority of these can interact with various transcription activators (hence the term "coactivator" that is used to describe the HATs). In metazoa, the most-studied HAT is the global transcriptional coactivator p300 and the closely related protein CBP; this large protein (ca. 2400 amino acids) contains at least four distinct interaction interfaces that allow it to be targeted by an extraordinary variety of transcriptional regulators, including proteins that regulate the cell cycle (such as c-jun and c-Fos), cell differentiation (MyoD), cell-cycle checkpoints (p53), and the nuclear hormone receptors. Most importantly, as discovered in the laboratory of Y. Nakatani in 1996, from an enzymatic standpoint CBP is a HAT capable of hyperacetylating all four core histones in solution. The ubiquity of CBP/p300's involvement in transcriptional regulatory pathways *in vivo* is powerful evidence to the pervasive use of targeted chromatin remodeling to effect gene control. In humans, mutations in the gene for CBP cause Rubinstein–Taybi syndrome—a multisymptomatic disorder characterized by mental retardation and a complex pattern of profound developmental abnormalities.

A well-characterized group of transcription factors that use HAT targeting to effect transcriptional activation are the nuclear hormone receptors—these recruit such HATs as SRC-1 and ACTR. Thus, HATs are an integral component of signal transduction pathways involving major regulators of mammalian physiology—glucose metabolism (glucocorticoids), ovulation (progesterone), development of secondary sexual characteristics (estradiol and testosterone), bone morphogenesis (vitamin D), brain development and skeletal maturation (thyroid hormone), and others. Point mutations in the gene for thyroid hormone receptor (TR) lead to the clinical disorder of resistance to thyroid hormone—patients present with goiter, stunted growth, and attention deficit hyperactivity disorder. Biochemical analysis in V. K. Chatterjee's lab revealed that the mutations impair the capacity of TR to recruit HAT coactivators and lower—or abolish, in some cases—its ability to activate transcription. These observations present very strong evidence that HAT targeting by specific transcriptional activators occurs *in vivo* and is relevant to genomic control.

The reaction opposite to that effected by HATs is catalyzed by histone deacetylases (HDACs). The history of their discovery begins, once again, in the mid-1980s, when a screen in budding yeast performed in the laboratory of R. Gaber identified a number of loci that reverse a potassium transport deficiency (=RPD) by virtue of the fact that mutations in those loci lead to the upregulation of potassium transporters. One such locus, *RPD3*, was subjected to additional genetic analysis and found to be involved in the control of transcription of a number of yeast genes. In 1996, the laboratory of S. Schreiber used an ingenious purification strategy to isolate a mammalian HDAC: these researchers reasoned that a competitive inhibitor could be used as bait for an HDAC (such a compound binds to the catalytic site of an enzyme but prevents catalysis due to its structural difference from the enzyme's *bona fide* substrate). Thus, the peptide trapoxin was used to prepare an affinity matrix through which crude mammalian cell extract was passed; two polypeptides were found to associate with the matrix—and peptide microsequence analysis of one of them revealed its close sequence similarity to budding yeast Rpd3p. Because activity assays indicated this newly purified protein to have HDAC activity, these data were accepted as the first evidence that an HDAC in mammals may be involved in transcriptional control. Subsequent analysis of mammalian RPD3 (also known as HDAC1) confirmed that notion.

A dedicated effort to clone additional HDACs from eukaryotic genomes revealed at least eight distinct genes—it is very likely that more will be identified once the emerging sequences of metazoan genomes, nematode, insect, rhodent, and primate, are analyzed in sufficient detail. Several general observations can be made at this time, however. From a functional standpoint, many of the HDACs have been functionally connected to transcriptional repression pathways; two examples are informative.

The first one involves one of the strongest mechanisms for effecting transcriptional repression currently known: DNA methylation. Genomes of higher vertebrates contain unusually low amounts of the dinucleotide CpG (theoretical predictions suggest each dinucleotide should account for $1/4^2 = 0.0625$, i.e., ca. 6% of the genome; CpG

is significantly less abundant). In fact, the bulk of CpG dinucleotides occur within short (ca. 100–500 base pairs) stretches of the genome designated “CpG” islands—these occur most commonly in the promoters of genes. In addition to this peculiar distribution, the cytosine within CpG dinucleotides is frequently covalently modified with a methyl group on position 5 (to yield 5-methylcytosine, 5-mC). There is very strong evidence that DNA methylation is a mechanism for targeting the transcriptional repression machinery to particular DNA sequences: for example, in our genomes, the bulk of genomic parasites (i.e., transposons and retroviruses) are kept transcriptionally quiescent by methylation. Experiments indicate that the demethylation of repetitive DNA can have severe consequences for genomic stability: for example, retroelements begin uncontrolled proliferation in the genomes of interspecific hybrids in wallabies, and mutations in the enzyme that effects DNA methylation (DNA methyltransferase) in humans leads to marked chromosome instability due to pericentric heterochromatin expansion (this genetic disorder—ICF syndrome—is characterized by skeletal abnormalities and mental retardation). In addition, a very common feature of neoplasia in humans is the aberrant silencing of genes required for cell cycle arrest (such as genes for cyclin-dependent kinase inhibitors); this is effected by hypermethylating the CpG islands in the promoters of these genes.

A functional connection between DNA methylation and the targeting of histone deacetylase emerged from studies on transcriptional regulators discovered in A. Bird’s lab that bind methylated DNA selectively (i.e., these proteins bind 5mCpG but do not bind CpG) and appear to be transcriptional repressors. Subsequent biochemical analysis in A. Bird and A. Wolffe’s lab revealed that one such protein—MeCP2—associates with histone deacetylase and that an HDAC inhibitor prevents transcriptional repression driven by MeCP2. These observations illuminated the fact that hypermethylated DNA sequences reside in deacetylated chromatin *in vivo*: thus, a simple model emerges according to which methylated DNA recruits specific proteins that recognize it selectively, and target HDAC to remodel chromatin adjacent to their binding site into a repressive, deacetylated conformation. The profound relevance of this pathway to genomic control *in vivo* can be seen from clinical and genetic evidence on human patients with mutations in the gene for MeCP2; these individuals develop Rett syndrome—a progressive, debilitating neurological disorder.

The second example regarding HDAC targeting in transcriptional repression involves certain members of the nuclear hormone receptor (NHR) superfamily. Some of these proteins—GR, for example—are constitutively cytoplasmic and translocate to their site of action, the nucleus, only in the presence of ligand. In contrast, other NHRs, including the retinoic acid receptor and the thyroid hormone receptor, reside in the nucleus irrespective of hormone. Their residence in the nucleus is not a passive one—both TR and RAR remain bound to target gene promoters in the absence of hormone and act as potent transcriptional repressors. The exact mechanism whereby these proteins effect repression remained elusive until work in 1995 from D. Moore, R. Evans, and M. Rosenfeld identified two large (ca. 270 kDa) related polypeptides called N-CoR and SMRT. Subsequent work from these labs, and also those of M. Lazar and A. Wolffe revealed unliganded TR and RAR can associate with HDAC via these corepressors, and that HDAC inhibitors will impair the repressive action by these regulators (the actual enzyme recruited by these proteins is HDAC3). Whether the targeting of HDAC leads to chromatin deacetylation of any vertebrate genes is not known. In budding yeast, work from the laboratories of M. Grunstein and K. Struhl showed (via the application of ChIP) that the transcriptional repressor Ume6p acts to silence target genes via the recruitment of the HDAC Rpd3p, and that such recruitment leads to histone tail deacetylation over target gene promoters.

The action of HDACs transcriptional repression pathways is not an academic issue, since HDAC inhibitors have long been known to be very potent cytostatic and differentiating agents. Thus, for example, a number of cell lines of cancerous origin can be driven to cease proliferation via the application of such HDAC inhibitors as trichostatin A or sodium butyrate. In addition, oncoproteins that are produced as a result of chromosomal translocations in leukemias are known to depend on HDAC targeting for action; compounds that force a release of HDAC from these chimeric proteins are successfully used in clinical practice to treat certain forms of leukemia.

An interesting property of HDACs is their tendency to occur in large, multisubunit complexes. For example, as originally discovered in the laboratory of A. Wolffe, and subsequently by the laboratories of S. Schreiber and D. Reinberg, the predominant biochemical form of HDAC1 in our cells is the Mi-2/NRD complex; its most distinguishing feature is that it contains a histone deacetylase, a SWI/SNF family ATPase, and a protein that can selectively bind to methylated DNA. An understanding of the functional relevance of these associations awaits futher experimentation.

V. CHROMATIN AND TRANSCRIPTION: A SYNTHESIS

Scientists that devoted their careers to the study of chromatin are understandably pleased with the progress of the

past decade: not only is our structural understanding of chromatin (at least on its elementary level) more profound than it has ever been but also a large number of protein complexes has been discovered that are intimately involved in transcriptional control and that can remodel and modify chromatin structure. These developments are somewhat of a mixed blessing, however, because the abundance of these protein factors, and our less-than-complete understanding of their *in vivo* function complicates attempts to depict gene activation and repression as a simple, linear sequence of events (e.g., "protein binds to DNA, protein recruits RNA polymerase, RNA polymerase synthesizes mRNA").

Two major questions in the field currently lack a comprehensive answer: (i) what are the *in vivo* structural consequences of chromatin remodeling and modification in terms of the behavior of the transcriptional machinery?; (ii) what is the relative contribution that chromatin modification/disruption, and non-chromatin-based regulatory pathways make to gene activation and repression *in vivo*? While comprehensive answers are currently lacking, an attempt at a synthesis of the current data can be made, however, based on experiments on the budding yeast HO endonuclease gene (K. Nasmyth), MMTV LTR (G. Hager), mouse serum albumin gene enhancer (K. Zaret), the *Xenopus* TRβA gene (A. Wolffe), and the Gal4-VP16 activator (A. Belmont). This hypothetical scenario for gene activation is presented in the form of a numbered list, but we emphasize that the sequence of some steps may be changed on specific promoters, and that some steps may be omitted altogether.

1. A transcriptional regulator accesses its binding site within a target promoter assembled into a mature nucleosomal array.
2. The chromatin-bound regulator targets an ATP-dependent chromatin remodeling engine such as SWI/SNF; this targeting leads to the localized remodeling of the histone DNA contacts and the generation of a DNAse I hypersensitive site.
3. This remodeling allows access to the promoter of other nonhistone factors; in concert with the "pioneer factor," these target HAT-containing complexes; their action may promote the large-scale unfolding of chromatin, as well as exert some local effect.
4. Some of the transcriptional regulators bound to the promoter use other adapter complexes to promote the assembly of the preinitiation complex and the targeting of RNA polymerase.
5. A number of complex events lead stalled RNA polymerase to begin productive transcriptional elongation (which may be facilitated by hyperacetylation).

It is very clear that transcriptional control harbors pathways and mechanisms that "are not dreamt of" in our current paradigms. The integration of chromatin infrastructure into gene expression regulatory pathways, however, is fairly certain to remain at the core of our notions of genome control.

SEE ALSO THE FOLLOWING ARTICLES

Cell Death (Apoptosis) • Gene Expression, Regulation of • Immunology-Autoimmunity • Protein Folding • Protein Structure • Ribozymes • Translation of RNA to Protein

BIBLIOGRAPHY

Bird, A. P., and Wolffe, A. P. (1999). "Methylation-induced repression—belts, braces, and chromatin," *Cell* **99,** 451–454.

Elgin, S. C. R. (ed.). (1995). "Chromatin Structure and Gene Expression," IRL Press, Oxford.

Grunstein, M. (1997). "Histone acetylation in chromatin structure and transcription," *Nature* **389,** 349–352.

Kingston, R. E., and Narlikar, G. J. (1999). "ATP-dependent remodeling and acetylation as regulators of chromatin fluidity," *Genes Dev*. **13,** 2339–2352.

Lemon, B., and Tjian, R. (2000). "Orchestrated response: A symphony of transcription factors for gene control," *Genes Dev.* **14,** 2551–2569.

Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997). "Crystal structure of the nucleosome core particle at 2.8 A resolution," *Nature* **389,** 251–260.

Ng, H. H., and Bird, A. (2000). "Histone deacetylases: Silencers for hire," *Trends Biochem Sci.* **25,** 121–126.

Robertson, K. R., and Wolffe, A. P. (2000). "DNA methylation in health and disease," *Nature Rev. Genet.* **1,** 11–19.

Sterner, D. E., and Berger, S. L. (2000). "Acetylation of histones and transcription-related factors," *Microbiol. Mol. Biol. Rev.* **64,** 435–459.

Struhl, K. (1998). "Histone acetylation and transcriptional regulatory mechanisms," *Genes Dev.* **12,** 599–606.

Sudarsanam, P., and Winston, F. (2000). "The Swi/Snf family: Nucleosome-remodeling complexes and transcriptional control," *Trends Genet.* **16,** 345–351.

Urnov, F. D., and Wolffe, A. P. (2001). "A necessary food; Nuclear hormone receptors and their chromatin templates," *Mol. Endo.* **15,** 1–16.

Wolffe, A. P. (1998). "Chromatin Structure and Function," Academic Press, San Diego, CA.

Wolffe, A. P., and Hayes, J. J. (1999). "Chromatin disruption and modification," *Nucleic Acids Res.* **27,** 711–720.

Wolffe, A. P., and Matzke, M. A. (1999). "Epigenetics: Regulation through repression," *Science* **286,** 481–486.

DNA Testing in Forensic Science

Moses S. Schanfield
Monroe County Public Safety Laboratory

I. What Is DNA?
II. Extraction of DNA
III. Quantitation of DNA
IV. Current Forensic DNA Testing
V. STR**s** Used Forensically
VI. PCR Setup and Detection
VII. Usefulness in Detecting Mixtures
VIII. Individualization

GLOSSARY

Alleles Alternate forms of an inherited trait. Such as the "A," "B," or "O" alleles of the ABO blood group. Simple polymorphisms will have as few as two alleles, while hypervariable polymorphisms will have five or more alleles, such that the majority of the population is heterozygous.

Amplified fragment length polymorphisms (AFLPs or AmpFLPs) Polymorphism in length of DNA segments detected using the PCR process to amplify a specific segment of DNA, followed by separation using electrophoresis. These are normally Variable Number Tandem Repeat (VNTR) regions that can have different size repeats. AFLPs are classified by the size of the repeat, repeats consisting of 10 or more bases are referred to as large tandem repeat (LTR) AFLP, while those containing seven or fewer base pairs per repeat are referred to as short tandem repeats (STRs).

Gene A sequence of DNA which codes for the production of a specific protein or part of a specific protein, such as a glactosyl transferase (the B allele of the ABO blood group).

Genotype The genetic type of a person determined either by the fact that they have two different alleles at a locus, or that the individuals parents have been tested to determine what alleles the person has genetically inherited.

Heterozygotes An individual with two different alleles at a given locus.

Homozygotes An individual with two alleles that are the same.

Locus The place (location) where a specific gene resides in the human genome. The names of genetic regions have been standardardized by the International System of Gene Nomenclature (Shows *et al.*, 1987). Inherited traits that code for genes usually have a name that reflects its biological function, such as THO1 for the

"Mentioning of commercial products is not an endorsement by The Monroe County Public Safety Laboratory."

first tyrosine hydroxylase locus. Some identified inherited traits do not code for proteins, but are regions that show variation (polymorphism) in length or sequence of DNA. These areas are referred to as "DNA loci," such as D1S7, which represents and DNA inherited trait (**D**), located on chromosome **1**, "**S**" indicates it is a single copy region and "**7**" indicates it was the seventh polymorphism found on that chromosome.

Phenotype The observed results of a genetic test. If a person has two different alleles (e.g., is heterozygous), the phenotype and the genotype are the same. However, if a person only has one allele detected, without testing the parents we cannot be certain that the person has two alleles the same (i.e., homozygous) since the person could have two alleles the same, or one detected allele and one undetected allele, for what ever reason. In the case of RFLP loci this is usually referred to as a single band pattern. For AFLP- and sequence-based PCR-based systems this is referred to as a homozygous phenotype. In general since we do not know the genetic type of homozygous individuals it is better to refer to homozygous and heterozygous phenotypes, unless the genetic type of the individual is determined by pedigree analysis.

Polymorphism Genetically inherited variation with two or more forms, the least common of which occurs at a frequency of greater than 1%.

Restriction fragment length polymorphism (RFLP) Polymorphism in length of DNA segments detected using a restriction enzyme, followed by separation using electrophoresis.

Variable number tandem repeat (VNTR) regions of inherited variation that consist of alleles which contain different numbers of repeating segments. It is easiest to think of them as freight trains with different numbers of box cars. Different loci will have different numbers of repeats. Though RFLP loci are also VNTR loci the number of repeats is generally not known because of the detection technology.

FORENSIC SCIENCE is applied science. That is to say that the scientific methodologies used in forensic science were developed by biologists, chemists, and geneticists and then taken over by forensic scientists to help them solve problems. The first inherited trait to be used in forensic testing was the ABO blood group on red blood cells and secreted blood group substance found in saliva and other body fluids of individuals called "Secretors." Research in the 1950s, 1960s, and 1970s identified proteins that had genetic variation or were "polymorphic." Some of these were enzymes such as acid phosphatase [ACP, referred to as erythrocyte acid phosphatase (EAP) by forensic scientists], esterase D (ESD), phosphoglucomutase (PGM1) and some transport and functional proteins such as group specific component [GC, now known as Vitamin D binding globulin [VDBG]], haptoglobin (HP), the immunoglobulin allotypes (GM and KM), and transferrin (TF) were used forensically. These markers were used in the late 1970s and early 1980s by forensic science in the United States and abroad to individualize blood stains. Although, some of these markers did not last long in bloodstains they made it possible to often individualize bloodstains with greater than a 99% certainty. Semen stains, saliva, and urine had relatively few markers that could be detected, making it difficult to provide information in cases with these types of evidence.

In the mid-1970s two independent areas of research would change the future of forensic science. Research on bacterial enzymes that cut DNA at specific places led to the development of restriction fragment length polymorphisms (RFLPs). Initially these were only useful for the diagnosis of genetic diseases such as Sickle Cell Anemia caused by a mutation in the hemoglobin gene. In 1980 with the identification of the first hypervariable DNA polymorphism (D14S1), detected by restriction length polymorphism technology (RFLP), the door was opened to the possibility that DNA technology could be applied to forensic evidence. In the next several years, the search for new markers lead to the identification of many forensically useful markers, detected by RFLP, some of which are still used routinely in forensic DNA testing. In the meantime, hypervariable minisatellite regions which identified many genetic regions at one time (multilocus probes) were found. The term "DNA fingerprinting" was used to describe these bar code-like patterns. Although these regions proved to be highly informative for parentage testing, they did not have the sensitivity needed for forensic testing. Though multilocus probes were used for paternity testing and some forensic applications, they are rarely used at the present time. Using a battery of five to seven RFLP loci made it possible to individualize samples into the 100s of millions and billions. This means that the chance of any two unrelated individuals matching was very unlikely.

At the same time the revolution in RFLP was beginning in the mid 1970s, an early version of copying DNA using repair enzymes called polymerases was being explored. It would not be until the early 1980s when the modern Polymerase Chain Reaction (PCR) tests was developed. The role of the polymerase enzymes in copying DNA had been known since the 1970s. The use of high temperature Taq polymerase allowed for the automation of thermal cycling and the introduction of modern PCR. Tests were developed to identify human leukocyte antigens (HLA) for the transplantation community. The first marker that they developed was to the HLA region called DQα (now called

DQA1). This marker looked at a genetic marker that had originally been tested at the protein level at the DNA level. This type of marker looked at DNA-sequence-based differences. This became the first PCR based tested to be used forensically. Other sequence-based tests were developed but they did not provide the same level of identification produced by RFLP based testing. Using the available sequence based tests only allowed individualization in the 100s to 100,000s.

The FBI established a standardized system for publicly funded crime laboratories in the United States. Similar work was going on in Canada, the United Kingdom, and Europe. The introduction of DNA restriction fragment length polymorphism (RFLP) technology revolutionized the field of forensic identity testing. This is especially true for the area of sexual assault evidence that historically has been an area of limited information. With DNA-based testing sperm DNA could be separated from the victims type allowing for the first-time regular direct testing of the sperm donor. Though RFLP technology has been a tremendous aid, it has several problems. It is expensive to implement, labor intensive, expensive to test, and is limited by both quantity and quality of DNA obtained. Further, because the process involved measuring the movement of bands and not directly the DNA product there were many statistical problems with representing the data. The technical feasibility of amplifying specific segments of DNA using the polymerase chain reaction (PCR) had the potential to overcome the shortcomings of RFLP technology.

PCR-based technology is much less expensive to implement, since it does not require a laboratory capable of handling radioactive isotopes. It has higher through put since each worker can do more cases in the same amount of time. PCR by its nature works with smaller amounts of DNA and with DNA that has been environmentally abused. Finally, since the DNA product can be identified as different alternative forms the statistical manipulation of data was similar to that for other genetic polymorphisms.

In 1989 at the same time that the RFLP-based testing was being converted to PCR-based testing creating the first amplified fragment length polymorphism or AFLP, new polymorphisms were being found directly using PCR. The converted RFLP loci which consisted of different numbers of repeated segments, much like freight trains with different numbers of box cars were referred to as “variable number of tandem repeat” or VNTR regions or loci. These repeats consisted of 15 to 70 base pairs and were referred to at ”large tandem repeat loci or LTRs. The new regions being found had much smaller repeats consisting of two, three, four, or five bases in a repeat unit. These new markers were called “short tandem repeats” or STRs for short. The four base pair or tetra nucleotide repeats became the genetic markers used to map the human genome. With this large number of markers available it became possible to pick sets of these markers to make highly discriminatory multiplexes (multiple regions amplified at the same time). The use of fluorescent detection of DNA fragments on automated DNA sequencers from the human genome project made it possible to create fluorescent multiplexes with automated detection. These methodologies are now the methods of choice for use in the forensic testing of DNA samples.

I. WHAT IS DNA?

DNA stands for deoxyribonucleic acid. It is the biological blueprint of life. DNA is made up of a double-stranded structure consisting of sugar (deoxyribose) and phosphate back bone, cross linked with two types of nucleic acids referred to as purines (adenine and guanine) and pyrimidines (thymine and cytosine) (Fig. 1). The cross linking nucleic acids always pair a purine with a pyrimidine, such that adenine always pairs with thymine and guanine always pairs with cytosine.

DNA can be found in several areas of a cell. The majority of DNA is located in the nucleus of cells (Fig. 2) organized in the form of chromosomes (22 pairs of autosomes and a set of sex chromosomes (X and Y)). Each nucleated cell normally has 46 chromosomes that represent the contribution from both parents. In the formation of gametes (eggs and sperm) one chromosome of each pair is randomly separated and placed in the gamete. The separation of chromosomes is referred to as segregation. The transmission of half of our chromosomes to our children in the form of gametes is the basis of Mendelian inheritance. This DNA is referred to as nuclear or genomic DNA. With the exception of identical twins, no two people share the same genomic DNA sequence.

Another source of DNA is found in the mitochondria in the cytoplasm of cells (Fig. 2). Unlike nuclear DNA, which only has two copies of each genetic region, mitochondrial DNA is involved in energy production within the cell and can have between 100 and 10,000 copies per cell. Structurally, instead of a linear arrangement of DNA within chromosomes, mitochondrial DNA has a circular structure. Mitochondrial DNA is inherited from the mother because it is found in the cytoplasm which comes from the egg (ova).

A. Where Is DNA Found?

Nuclear or genomic DNA is found in all nucleated cells as well as in the reproductive cells (eggs and sperm). The amount of DNA we can expect to find in different cells and types of evidence are found in Table I. DNA has been

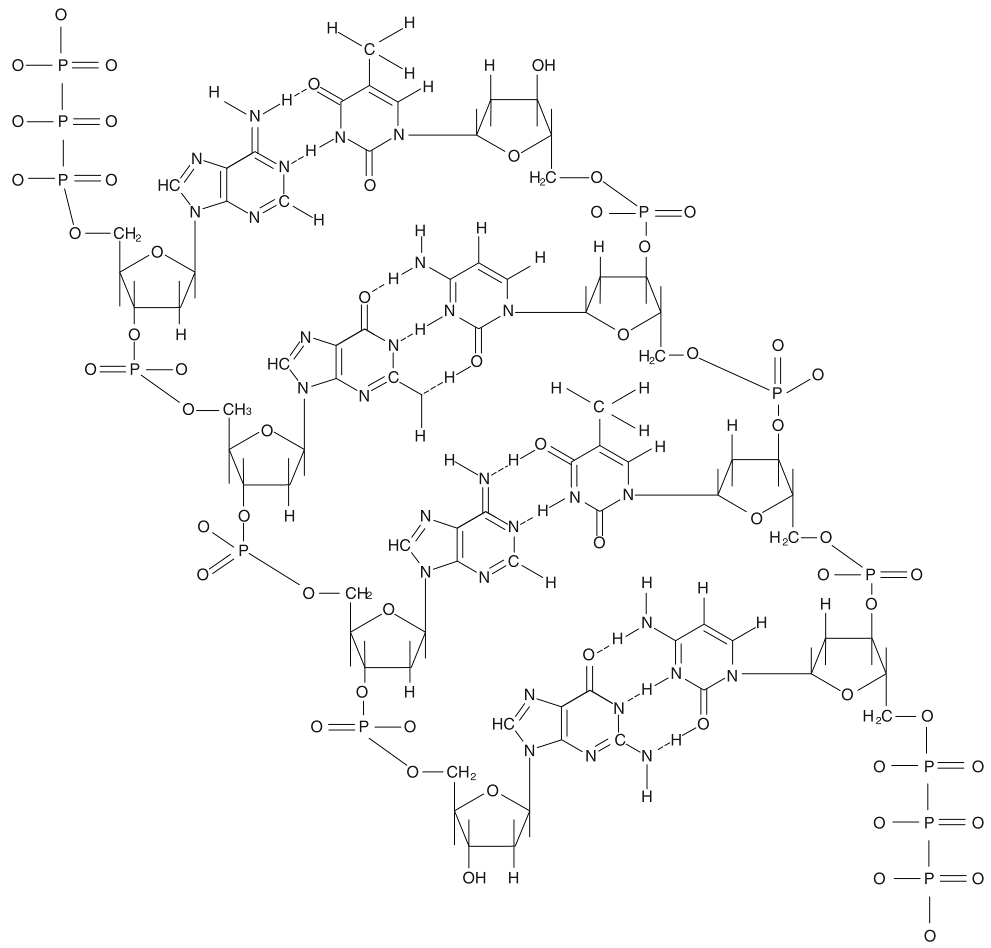

FIGURE 1 Molecular structure of DNA. From top to bottom: Adenine–Thymine, Guanine–Cytosine, Adenine–Thymine and Guanine–Cytosine. (From Schanfield, M. S. (2000). Deoxyribonucleic Acid/Basic Principles. *In* "Encyclopedia of Forensic Sciences" (Siegel, J. A., Saukko, P. J., and Knupfer, G. C., eds.), Academic Press, London, p. 481.)

successfully obtained from blood and bloodstains, vaginal and anal swabs, oral swabs, well-worn clothing, bone, teeth, most organs, and to some extent urine. It is less likely to obtain DNA from some types of evidence than others. Blood or semen stains on soil and leather are historically not good sources of evidenciary DNA. Saliva, per se, has few nucleated cells, but, beer and wine bottles, drinking glasses, beer cans, soda cans, cigarettes, stamps and envelope flaps have all been found to provide varying amounts of DNA.

B. How Much DNA Is Needed for Forensic Testing?

The amount of DNA needed to perform testing depends on the technology used. RFLP technology usually needs at least 50 ng of intact high-molecular-weight DNA. In contrast PCR-based testing can use as little as 250 pg. Most PCR based tests are set up to use between 1 and 10 ng of genomic DNA.

C. Destruction of DNA

Biological materials are going to be affected by their environment. Enzymes lose activity over time and type of storage conditions. DNA has been found to be relatively robust when it is in the form of dry stains. Initial environmental studies indicated some of the limitations of DNA based on the material it is deposited upon and the environmental conditions. Environmental insult to DNA does not change the results of testing, you will either obtain results, or if the DNA has been too badly affected by the environment (i.e., the DNA is degraded) you do not get RFLP results. One report on the success rate of obtaining RFLP results and noted that depending on the substrate or condition of the stain, results were obtained

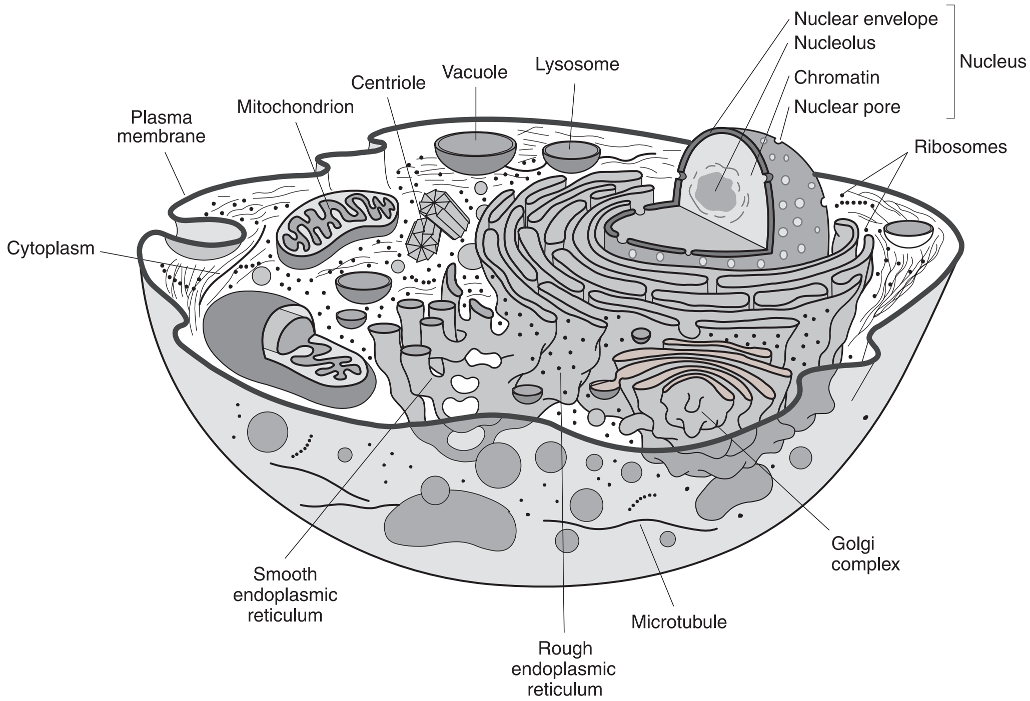

FIGURE 2 A generalized eukaryotic cell showing the organization and distribution of organelles as they would appear in transmission electron microscope. The type, number, and distribution of organelles are related to cell function. (From Schanfield, M. S. (2000). Deoxyribonucleic Acid/Basic Principles. *In* "Encyclopedia of Forensic Sciences" (Siegel, J. A., Saukko, P. J., and Knupfer, G. C., eds.), Academic Press, London, p. 481.)

between 0 (carpet stains or putrefied samples) and 61.5% (scrapped dried stains) of the time with and average of 52% for the 100 items of evidence tested.. Thus, the material DNA is deposited on and the degree of further insult can markedly affect the ability to obtain RFLP DNA results.

All of the published studies on environmental insult were done on prepared dried stains. Since biological fluids are liquid the effects of ultraviolet radiation on liquid DNA have been evaluated. The results of exposing 100-μl samples of a standard DNA solution to fluorescent light in the laboratory, a UV germicidal light (254 nm), midday sunlight in January, and early sunset light in January in 15-min increments, up to 1 hr are presented in Fig. 3. There is a linear decrease in high-molecular-weight DNA with the UV germicidal light, such that after an hour about 96% of the high-molecular-weight DNA has been lost. Even in the weak midday light in January, over 60% of the high-molecular-weight DNA was lost. In contrast, the fluorescent lighting in the laboratory and the after sunset light had

TABLE I DNA Content of Various Tissues

1 sperm	3 pg
1 cell	6 pg
1 shed hair	1 ng[a]
1 plucked hair	300 ng[b]
1 drop of blood	1,500 ng

[a] The success rate for PCR on shed hairs is 30 to 50%, so this average is optimistic.

[b] There is a great deal of variation among hair roots. Fine blond hair will tend to have much less DNA, while course dark hair with large roots more.

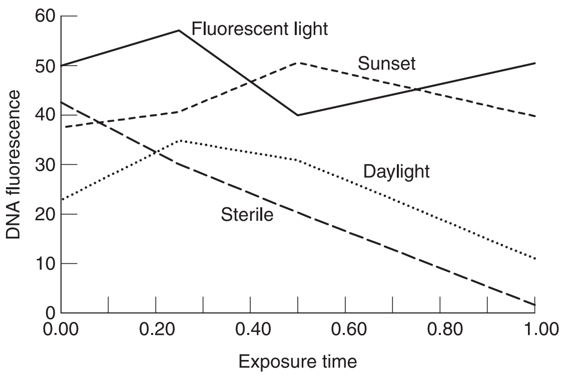

FIGURE 3 Plot of DNA concentration (froresecence) over time after exposure to different light sources. (From Schanfield, M. S. (2000). Deoxyribonucleic Acid/Basic Principles. *In* "Encyclopedia of Forensic Sciences" (Siegel, J. A., Saukko, P. J., and Knupfer, G. C., eds.), Academic Press, London, p. 482.)

no effect on the amount of high-molecular-weight DNA. This was not a rigorous experiment, but the effects are dramatic enough to demonstrate the effect of ultra violet light exposure to DNA before stains dry.

II. EXTRACTION OF DNA

As stated previously DNA exists inside of cells. Because most evidence is in the form of dry stains, the DNA must be removed from the stain before it can be tested. The process of removing DNA from the cells on the evidence and dissolving it is referred to as extraction. There are several procedures available for removing DNA from evidence so that it can be used. They are referred to as either "organic" extraction or "nonorganic" extraction based on the nature of the chemicals used. Further, there are two special types of extraction. The first, called differential extraction, was developed for sexual assault evidence to separate the cells that come from the victim (epithelial cells from the vagina, rectum, or mouth) from those of the perpetrator (male sperm cells). The second method is a specialized "nonorganic" extraction using Chelex beads. Chelex beads can only be used when PCR-based DNA testing is going to be used. The basic DNA extraction procedures, whether organic or nonorganic, can be adapted for special circumstances such as hair or tissue.

A. Chloroform–Phenol Extraction

This is the oldest procedure available for extracting DNA from blood and it has been extended to include hair, tissue, and semen stains. The basic procedure consists of opening up cells with a buffer and an enzyme, usually Protease K, and then denaturing and separating the proteins from the DNA. The latter part is done using a mixture of chloroform (24:1 chloroform:isoamyl alcohol) and phenol (buffered). The phenol–chloroform mixture denatures proteins liberated by the first stage. The major disadvantage of this procedure is the fact that phenol–chloroform is a hazardous waste and could theoretically pose a risk to pregnant employees. A modern protocol for phenol–chloroform extraction of various types of evidence can be found in the literature.

B. Nonorganic Extraction

In nonorganic extraction the hazardous phenol–chloroform protein denaturation step was replaced by a salting out of proteins. This allowed for the same chemistry to be used for the initial phase of DNA extraction, and replacement of the hazardous elements of the procedure with a nonhazardous alternative. The salting-out procedure has several advantages over the phenol–chloroform extraction. The first is that instead of having two liquid phases (organic and nonorganic) that can occasionally trap the DNA in the wrong phase (organic) phase, by precipitating the proteins (e.g., the proteins become insoluble and become a solid), there are liquid and solid phases with the DNA only in the liquid phase (nonorganic). The second advantage is that the hazardous phenol–chloroform is replaced with a harmless salt solution. Comparison of the organic and nonorganic procedures for blood, and semen indicate that the nonorganic extraction is on the average as good or better than organic extraction, whether quantitated by yield gel or slot blot (Table II).

Either method of DNA extraction described earlier can be used for both RFLP- or PCR-based DNA testing. Organic DNA extraction is widely used in laboratories doing criminal casework while nonorganic DNA extraction is

TABLE II Comparison of Organic and Nonorganic Extraction of DNA from Blood and Semen Stains[a]

Quantitation method	Blood organic		Blood nonorganic	Semen organic		Semen nonorganic
Yield gel						
Mean	185 ng		258 ng	175 ng		207 ng
N	21		8	22		8
p		.054			.122	
Slot blot						
Mean	515 ng		908 ng	627 ng		1175 ng
N	22		8	27		8
p		.022			.008	

[a] Data taken from Tables 1 and 2 of Laber *et al.* (1992). Differences in means tested by Kruskall–Wallace nonparametric analysis of variance, H statistic with 1 df, uncorrected p values presented.

widely used in laboratories performing paternity testing, research, and diagnostics. On a worldwide basis nonorganic DNA extraction is the more prevalent. With the shift to PCR-based testing this choice in extraction is increasingly common.

C. Chelex Extraction

In 1991, a method of DNA extraction was described that was specifically aimed at the extraction of small amounts of dilute DNA for PCR-based testing using Chelex beads. The method is simple, relatively fast, and biohazard free. It is widely used by forensic laboratories doing PCR-based typing which has increased the number of laboratories using nonorganic, biohazard-free DNA extraction. The only limitations of Chelex extraction is that it produces a dilute solution of DNA that may need to be concentrated before it can be used with some of the newer high-resolution PCR-based typing systems and it cannot be used for RFLP testing.

III. QUANTITATION OF DNA

A. Yield-Gel Quantitation

Whether RFLP- or PCR-based testing is performed it is necessary to know how much DNA is present. One of the earliest methods of quantitating small amounts of DNA is the use of a yield gel. A small gel is made using a salt solution to carry electrical current and a supporting medium made of agarose (a complex carbohydrate made from seaweed). Much like gelatin, the agarose dissolves in water that is heated to near boiling and the liquid is cooled slightly and poured into a casting tray. A plastic comb or well former with rectangular teeth is placed in the liquid agarose. Once the agarose gels, the comb is removed leaving behind rectangular wells in the agarose gel. The DNA to be tested is mixed with loading buffer, and placed in the wells. Loading buffer is a mixture of a large amount of sugar and dye. The high concentration of sugars makes the mixture heavier than the salt solution so that the DNA sinks to the bottom of the well. The dye allows the migration of the DNA to be monitored. The agarose was melted in water containing salt. When electrical current is applied to the gel, electricity flows through the gel because of the salt and moves (migrates) from the negative electrode (cathode) toward the positive electrode (anode). Since all DNA has a negative charge, and was placed in the wells at the cathodal end of the gel, the negatively charged DNA will migrate out of the wells toward the positive end of the gel. If the DNA is broken into pieces that are different sizes, the smaller pieces will move through the gel faster than the larger pieces and will be separated based on size. This process of separating DNA using an electric current is called electrophoresis, which simply means separation (phoresis) by means of electricity (electro).

Since DNA is colorless it is not possible to see the DNA after it has been separated without the use of special dyes that bind to it. One of the earliest dyes used was ethidium bromide, which fluoresces pink when bound to double-stranded DNA and exposed to ultraviolet light. Figure 4 is an ethidium bromide stained yield gel. To quantitate the amount of DNA in the DNA extracts, a set of DNA quantitation standards are placed on the gel. By visual comparison of the unknown DNA to the known DNA the amount of DNA can be approximated. This test provides information about the relative amount of DNA and whether it is degraded (i.e., the DNA is broken down so that different-size pieces of DNA are present). It does not indicate if the DNA is human, however, since all DNA will fluoresce. Thus the DNA present may be bacterial as well as human DNA. For RFLP testing the total amount of DNA in the sample is the important determinant of how the samples migrate in the gel. Therefore yield-gel electrophoretic quantitation of DNA is an appropriate method. Yield-gel quantitation of DNA for RFLP testing was considered to be such an integral part of quality assurance that it was included in the National Institute of Standards, Standard Reference Material 2390, "DNA Profiling Standard."

As with the extraction of DNA using the organic method, ethidium bromide is potentially hazardous because the dye is associated with an increased cancer risk. Though ethidium bromide is still widely used for the identification of DNA it is currently being replaced by a new dye called Sybrr® green which is much less carcinogenic and can detect smaller amounts of DNA than ethidium bromide.

B. Slot-Blot Quantitation

In contrast to RFLP, for PCR-based testing, the amount of human DNA and not the total amount of DNA is an important determinant in how likely it will be to obtain results. A slot blot does not rely on electrophoresis to separate the DNA but rather on the ability of denatured (separated DNA strands) DNA to bind to homologous complementary sequences. The ability to quantitate human DNA requires sequences of DNA that are common in the human genome so that a single DNA sequence can recognize them and bind to them. The repeated DNA sequence called D17Z1 is the basis for all human DNA slot-blot quantitation systems. There are several of these procedures commercially available. In one of the most widely used tests, the quantitation requires that denatured DNA is applied to a membrane using a slotted plastic apparatus. The denatured DNA binds to the membrane. The

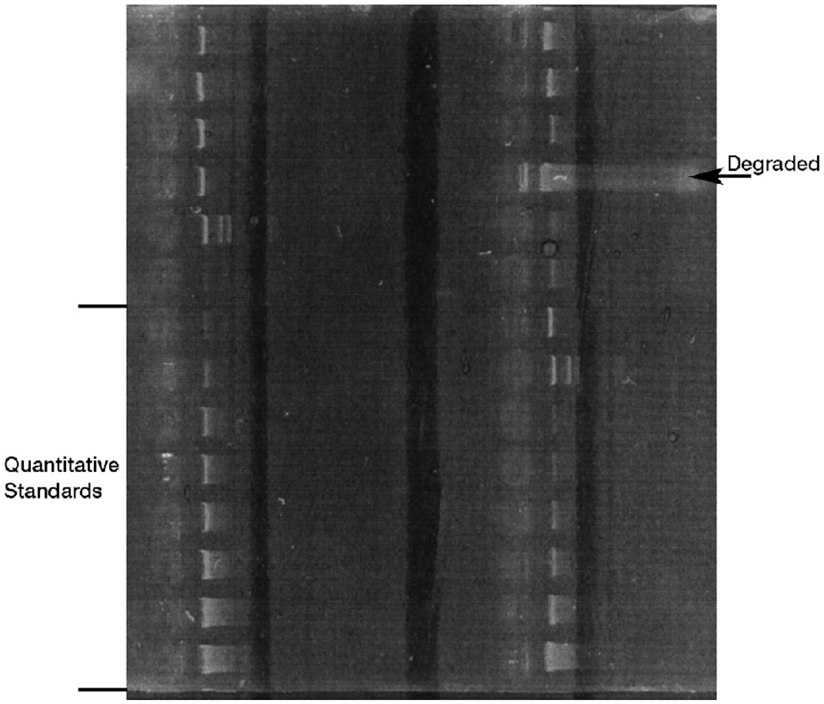

FIGURE 4 Ethidium bromide stained yield gel. Bottom left samples are quantitative standards. Other samples represent various samples of DNA. Upper right sample is degraded DNA. (From Schanfield, M. S. (2000). Deoxyribonucleic Acid/Basic Principles. *In* "Encyclopedia of Forensic Sciences" (Siegel, J. A., Saukko, P. J., and Knupfer, G. C., eds.), Academic Press, London, p. 484.)

membrane is exposed to a solution of denatured DNA fragments that recognizes a repeating sequence of human or primate DNA. Pieces of DNA that recognize a specific region of DNA are referred to as a "probe." The probe will bind to complementary DNA fragments stuck on the membrane. The probe has an indicator attached to it so that the binding of the DNA to the probe can be detected. The unbound probe is washed off and the probe is detected using either chemicals that change color (colorimetric detection) or chemicals that give off light (chemiluminscent detection). To be able to quantitate the amount of human DNA present, standards with different amounts of human DNA are also placed on the membrane so that it is possible to determine the approximate amount of DNA bound to the membrane by visual comparison to the known standards. More precise quantitation can be obtained by scanning the membrane with a scanning densitometer and determining the amount of color associated with each band. Most forensic laboratories use visual comparison.

IV. CURRENT FORENSIC DNA TESTING

At this point in time, at the beginning of the 21st century, forensic DNA testing has moved away from RFLP testing and is replacing sequence-based PCR strip technology with fluorescent STR-based testing. Thus it is important to understand the nature of PCR-based testing and the power it provides to forensic scientists.

A. Definition and Description of PCR

The Polymerase Chain Reaction (PCR) is based on biochemical processes within cells to repair damaged DNA and to make copies of the DNA as the cells replicate. In the repair mode, if a single strand of DNA is damaged, the damaged area is removed so that there is a single-stranded section of DNA with double-stranded sections at either end. The polymerase enzyme fills in the missing complementary DNA. In the copy mode an entire strand is copied during DNA replication. Figure 5 illustrates a polymerase enzyme copying a portion of a strand of DNA.

In a cell a specific gene is copied or translated from DNA to RNA because the polymerase has specific start and stop signals coded into the DNA. To copy a sequence of DNA *in vitro*, artificial start and stop signals are needed. These signals can only be made once the sequence of the region to be amplified is known. Once a sequence is known, the area to be copied or amplified can be defined by a

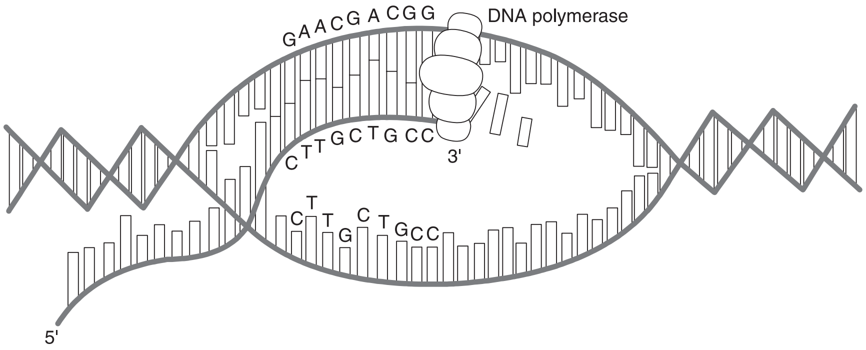

FIGURE 5 DNA polymerase copying one strand of a portion of double-stranded DNA. (From Schanfield, M. S. (2000). Deoxyribonucleic Acid/Polymerase Chain Reaction. *In* "Encyclopedia of Forensic Sciences" (Siegel, J. A., Saukko, P. J., and Knupfer, G. C., eds.), Academic Press, London, p. 516.)

unique sequence of DNA. For a primer to recognize a unique sequence in the genome it must be long enough for no other sequence to match it by chance. This can usually be achieved with a sequence of 20 to 25 nucleotides. These manufactured pieces of DNA are called "primers" and they are complementary to the start and stop areas defined earlier. The "forward primer" is complementary to the beginning sequence on one strand of DNA, usually called the positive strand. The "reverse primer" is complementary to the stop sequence on the opposite or negative strand of DNA.

B. Multiplexing PCR Reactions

One of the advantages of PCR is that more than one region can be amplified at a time. Although it is necessary to select carefully primers that cannot bind to each other, the only limitation is how many pairs of primers can be placed together is the ability to detect the amplified product.

C. PCR Process

To perform a PCR reaction several ingredients are needed. They include PCR reaction buffer, which is basically a salt solution at the right pH for the enzyme being used, the four nucleotides (DNA building blocks), primers, a thermostable DNA polymerase (Taq, Pfu, Vent, Replinase, etc), and template DNA. These reactants are placed in small plastic reaction tubes. The process consists of heating a solution of DNA to greater than 90°C. Double-stranded DNA comes apart or melts to form single-stranded DNA at this temperature. This is called the *denaturation* step. The solution is then cooled down to between 50 and 65°C the primers will bind to their complementary locations. This is called the *annealing* or probe hybridization step. Finally, the solution temperature is raised to 72°C at which point the polymerase makes a copy of the target DNA defined by the primers. This is called the *extension* step. This completes one cycle of the PCR process. To make enough copies of the target DNA to detect the process is repeated from 25 to 40 times. This is done using a device called a thermalcycler. The process is illustrated in Fig. 6. If the process were perfect 30 cycles would create over a billion copies of the original target DNA.

The heating and cooling of the tubes are done in an electromechanical device call a "thermalcycler," which in general, consists of an aluminum blocks with wells designed to fit the plastic PCR reaction tubes. The aluminum block has heating and cooling elements controlled by a microprocessor that can raise and lower the temperature of the block and the plastic PCR reaction tubes in the block. In the thermal cyclers that were first made, the plastic reaction tubes extended above the thermal block. This allowed cooling to take place above the reaction. The water in the reaction mixture would evaporate and condense at the top of the tube, changing the concentration of reactants and affecting the success of the amplification. To limit the evaporation, mineral oil was placed on top of the reaction mixture. New thermal cyclers have heated lids on top of the block to prevent or minimize evaporation. The microprocessor can store many sets of instructions such that different programs can be kept in the microprocessor to amplify different sequences of DNA.

D. Detection of PCR Products

There are many methods for detecting PCR products. Since large amounts of product are produced there is no need to use techniques such as radioactive detection, although it has been used in some clinical setting. In forensic testing, one of the advantages of PCR-based testing is that it does not require the use of hazardous materials to detect it. There is normally enough product so that if the PCR products are run on a yield gel and stained with ethidium bromide or Cyber green, there is normally enough DNA

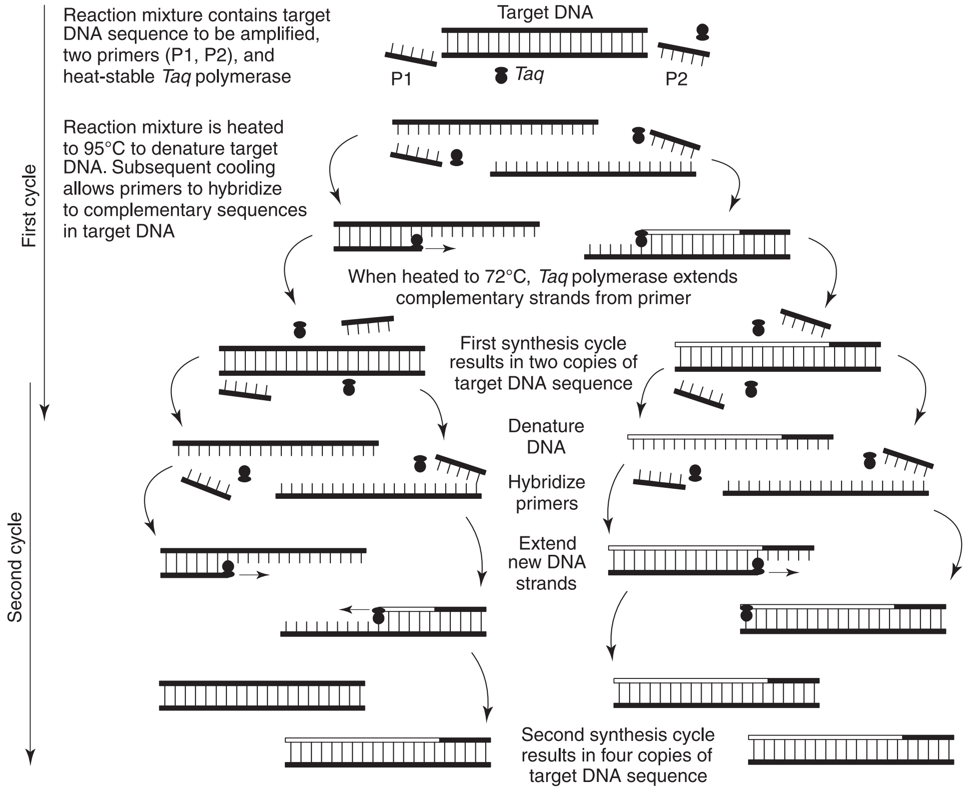

FIGURE 6 The PCR process. Courtesy of PE Cetus Instruments. (From Schanfield, M. S. (2000). Deoxyribonucleic Acid/Polymerase Chain Reaction. *In* "Encyclopedia of Forensic Sciences" (Siegel, J. A., Saukko, P. J., and Knupfer, G. C., eds.), Academic Press, London, p. 517.)

to be detected. This is a suggested method to verify if the PCR amplification was successful and that there is a PCR product to detect.

V. STRs USED FORENSICALLY

At this point in time with the large number of STR loci, a demand for a standardized panel in the United States, and a need for there to be at least some sharing of loci with forensic counterparts in Canada, England, and Europe, the Technical Working Group on DNA Analysis Methods (TWGDAM) implemented a multi-laboratory evaluation of those STR loci available in kits in the United States. The loci chosen would be the PCR-based core of a national sex offender file required under the 1994 DNA Identification Act. The national program is called Combined DNA Indexing System or CODIS for short.

The TWGDAM/CODIS loci were announced at the Promega DNA Identification Symposium in the fall of 1997 and at the American Academy of Forensic Science meeting in February 1998. The following loci were chosen to be part of what was originally called the CODIS 13 loci: CSF1PO, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, FGA, THO1, TPOX, and VWA03. These loci overlapped with the Forensic Science Services multiplexes and the Interpol multiplexes. The 13 loci can be obtained in two amplifications using Profiler Plus and Cofiler or in a single amplification using a kit in development called Identifiler from

1 and Powerplex 2 from Promega (Table II), or in a single reaction with Powerplex 16 by Promega.

A. Fluorescent Dyes

Before discussing the equipment used to detect fluorescent STRs some understanding of fluorescent dyes are necessary. Fluorescent dyes or minerals when subjected to light at one wave length, such as untraviolet light or black light, will give off colored light at a slightly different wavelength or color. A characteristic of fluorescent dyes or materials is that the compound is excited at one frequency of light, referred to as its absorption or excitation peak, and emits or gives off light a different frequency, referred to as its emission peak.

To label a PCR primer with a fluorescent dye, the dye is attached to the $5'$ end of the molecule. Since DNA is

translated from 5 to 3′ it is at the very end of the primer and should not affect the amplification or binding of the primer if made correctly. One of the oldest fluorescent dyes is fluoroscein. Many devices have been made that will detect fluoroscein, and it has been used extensively to label antibodies and other materials. Many of these dyes have been used for a long time and are in the public domain. Others have been developed for specific projects. The dyes used by PE Biosystems were originally proprietary and part of a patented four color DNA sequencing system (Blue, Green, Yellow, Red). These dyes are now becoming more readily available.

B. Fluorescent Detection Equipment

The equipment used to detect the product of fluorescently labeled STR tests fall into two categories. The first are devices that scan a gel after the DNA products have been separated by a process called electrophoreses. Examples of post electrophoresis scanners are the Hitachi FMBIO® Fluorescent Scanner, the Molecular Dynamics Fluorimager™ and the Beckman Genomics SC scanner. The Hitachi and Molecular Dynamics use a laser as a light with filters to identify the proper frequency and a CCD camera to capture the image. The Beckman Genomics SC uses a monochromatic Xenon light source and uses filters to detect the appropriate light for the CCD camera. The CCD camera scans back and forth over the gel as it is exposed to the light source and detects the various fluorescent colors using filters that change. This type of equipment has flexibility because different formats of electrophoresis gels can be used and scanned. The output is in the form of an electronic image with bands, that look much like a set of RFLP bands. Figure 7A is an example of actual images

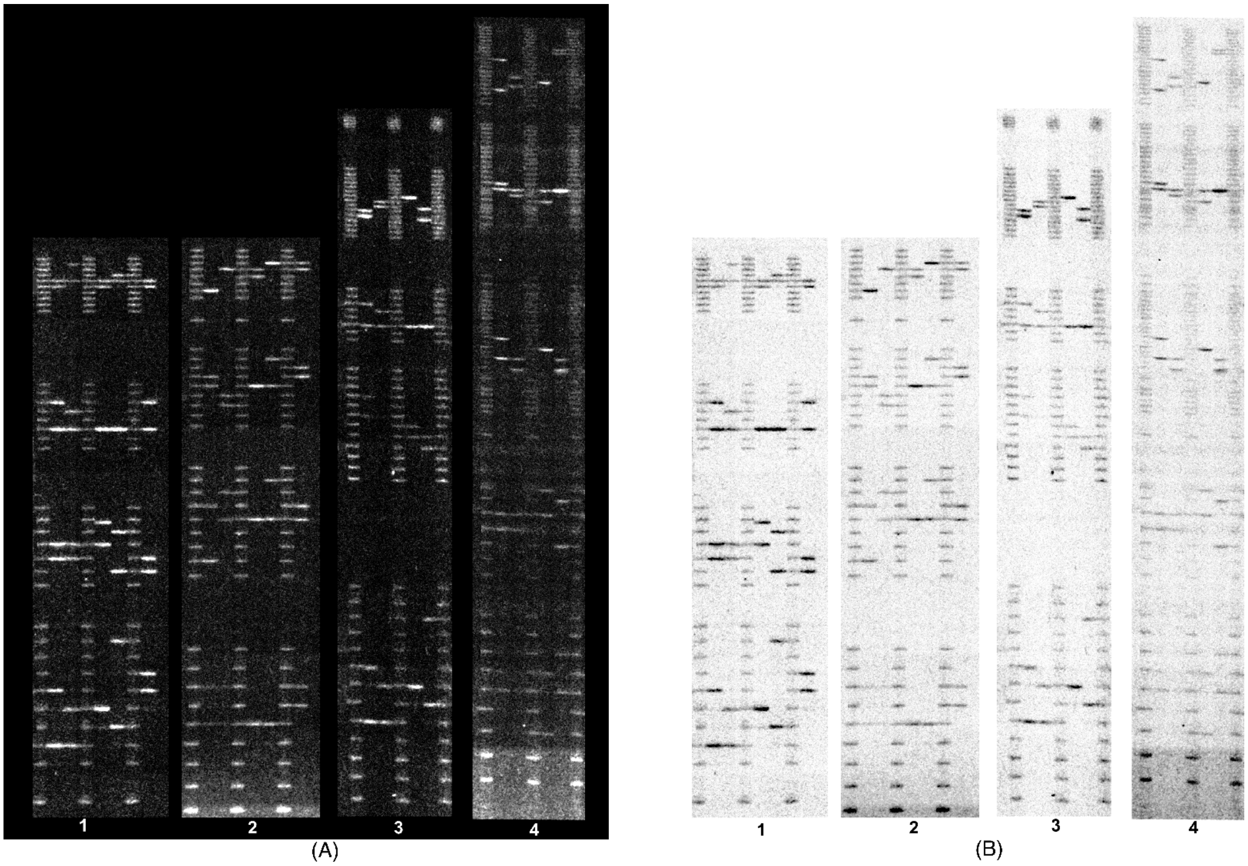

FIGURE 7 (A) is the original CCD images of Powerplex 1.1 (Panels 1 and 2) and Powerplex 2.1 (Panels 3 and 4). (B) is the reverse image of the same images. (B) Panel 1 is the TMR loci from top to bottom CSF1PO, TPOX, THO1, and VWA03. (B) Panel 2 is the fluoroscein loci, from top to bottom D16S539, D7S820, D13S317, and D5S818. (B) Panel 3 is the TMR loci from top to bottom FGA, TPOX, D8S1179, and VWA03. (B) Panel 4 is the fluoroscein loci from top to bottom Penta E, D18S51, D21S11, THO1, and D3S1358. The same samples are loaded between the ladders. The sample format is the TWGDAM format when an internal size standard is not used. The overlapping loci for sample confirmation are THO1, TPOX and VWA03. (From Schanfield, M. S. (2000). Deoxyribonucleic Acid/Polymerase Chain Reaction-Short Tandem Repeats. *In* "Encyclopedia of Forensic Sciences" (Siegel, J. A., Saukko, P. J., and Knupfer, G. C., eds.), Academic Press, London, p. 530.)

(light on dark) recorded by the CCD camera, and Fig. 7B is the reverse image (dark on light) that is reminiscent of an RFLP autoradiograph or lumigraph. It should be noted that the scanning time is added onto the electrophoresis time, with increased time for each color read.

The second type of imaging system is a real time system, in which the DNA fragments, after the bands have been resolved, pass beneath a light source scanner that recovers the spectrum of light from the different fluorophors. This is the ABI Prism® system from PE Biosystems. It includes the older Model 373 DNA Sequencer and 377 DNA Sequencer, which use slab acrylamide electrophoresis to separate the DNA fragments, and the 310 and 3100 Genetic Analyzers which use capillary electrophoresis to separate the DNA fragments. Capillary electrophoresis is a technology in which a fine glass capillary is filled with a proprietary separation polymer. The sample is pulled into the capillary by applying an electric current to it, then using high-voltage electrophoresis (12,000 V), and the DNA fragments are separated over the length of the column and move past a laser detector. The 377 can put approximately 60 samples on a gel at one time, and with modifications, 96. In contrast, the 310 CE system does one sample at a time, with a separation time of approximately 20 min. However, as this is automated, a cassette can be filled with samples for testing and left to run unattended. The 3100 uses 10 capillary tubes with higher throughput. The output of these devices is not a CCD image, but a series of electropherograms with a profile for each color scanned (nominally Blue, Green, Yellow, and Red). Since these are difficult to interpret the computer software provides decomposed single color graphs. Figure 8 contains an electropherogram of the single amplification Promega PowerPlex 16.2 System run on an ABI PRISM 310 Genetic Analyzer.

One of the extremely useful characteristics of fluorescent imaging devices is that the amount of light read by the detection device is quantified. The electropherograms produced by the real time scanners is quantitative. However, the CCD image can also be used as a scanning densitometer to determine the amount of light in each peak. Since there is one fluorescent molecule per band the amount of fluorescence is linear with the number of molecules in a band. This allows for many different types of analysis to be performed on the data generated. One of the more important of these is the ability to detect mixtures (see following).

VI. PCR SETUP AND DETECTION

The manufacturers of the kits have done forensic validations of the kits; however, each laboratory is responsible for the individualized validation required before testing. The guidelines for those validations for laboratories in the United States are governed by the DNA Advisory Board Guidelines (as of October 1, 1998) as implemented by ASCLD-LAB. Other areas of the world are regulated by other guidelines, unless they are also ASCLD-LAB accredited.

A. STR Detection

The major difference in the typing of the STR loci is the ability to include an internal size standard if the detection device used has multicolor capability. Under the TWGDAM Guidelines forensic samples are to be placed adjacent to an allele ladder, as seen in Fig. 7 (PCR–STR). Since the Beckman Genomyx SC only has two filters (fluoroscein and TMR) an internal ladder could not be used, so the adjacent ladder format is used. In this situation there is no special preparation for detection. When the four-color Hitachi FMBIO II Fluorescent Scanner, ABI Prism 377 or 310 is used, an internal standard is used to size the DNA fragements. As part of the electrophoresis setup a ROX ladder is added to PE Biosystems amplified products while a CRX ladder is added to Promega kits. (See Figure 8 for example.) Amplified products including the

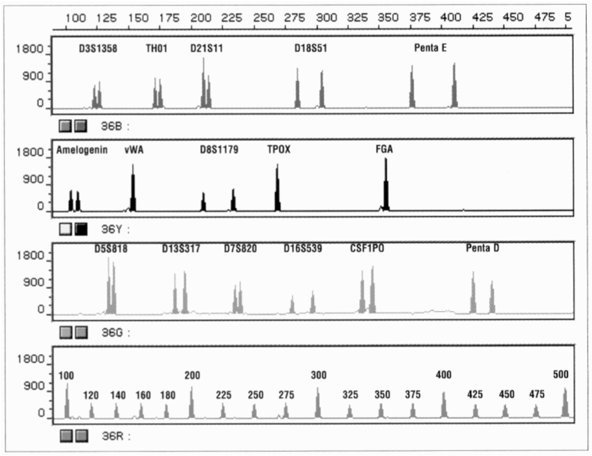

FIGURE 8 Electropherogram of a single DNA sample amplified using the 16-locus prototype Powerplex™ 16.2 System detected with the ABI PRISM® 310 Genetic Analyzer. All 16 loci were amplified in a single reaction and detected in a single capillary. The fluorescein-labeled loci (D3S1358, THO1, D21S11, D18S51, and Penta E) are displayed in blue, the TMR labeled loci (Amelogenin, VWA03, D8S1179, TPOX, and FGA) are displayed in black, and the loci labeled with a new dye (D5S818, D13S317, D7S820, D16S539, CSF1PO, and Penta D) are displayed in green. The fragments of the prototype ILS-500 size marker are labeled with CXR and are shown in red. (Taken from Promega promotional literature.) (From Schanfield, M. S. (2000). Deoxyribonucleic Acid/Polymerase Chain Reaction-Short Tandem Repeats. *In* "Encyclopedia of Forensic Sciences" (Siegel, J. A., Saukko, P. J., and Knupfer, G. C., eds.), Academic Press, London, p. 533.)

alllelic size ladders. The internal size standard is used to size all fragments within a lane by detector supplied software and assigns repeat numbers from the allele ladder sizings.

B. The Use of Internal Size Standards

It was previously demonstrated that within gel variation in DNA migration could be compensated for by placing a size ladder within the lane and measuring each fragment with the internal size standard. This allows for highly precise measurements of fragments. This is necessary since the electrophoresis systems used to detect the STR loci must have the capability of resolving differences between alleles as little as one base pair, to make sure that the fragment sizes can be accurately converted to repeat numbers. This would not be critical if all STR were regular. That is to say always four repeats for the tetra nucleotide STRs. However, this is not the case. Some STR loci have one or more common alleles that differ by only a single base pair (THO1, FGA, and D21S11). An example of this is seen in Fig. 7B, Panel 1, third sample from the left, in the third locus from the top.

VII. USEFULNESS IN DETECTING MIXTURES

One of the major problems in the analysis of forensic evidence is posed by samples containing biological material from more than one source. The large number of discrete alleles at multiple loci make STR multiplexes and excellent tool for identifying components of mixtures.

VIII. INDIVIDUALIZATION

In the United States, the FBI has started releasing reports indicating that biological material originated from a specific source, much as fingerprint examiners have done for many years. The FBI has decided that if the population frequency exceeds 1 in 230 billion the sample is individualized. Other laboratories have chosen high thresold levels such as 1 in 500 billion. Whatever the level chosen it is estimated that the average power of exclusion for these 13 CODIS loci exceeds 1 in a million billion, and though it is possible to obtain a frequency more common than that required for individualization it will occur infrequently.

SEE ALSO THE FOLLOWING ARTICLES

Hydrogen Bond • Mass Spectrometry in Forensic Science • Protein Structure • Spectroscopy in Forensic Science • Toxicology in Forensic Science • Translation of RNA to Protein

BIBLIOGRAPHY

Adams, D. E., Presley, L. A., and Baumstark, A. L., *et al.* (1991). "Deoxyribonucleic Acid (DNA) analysis by restriction fragment length polymorphism of blood and other body fluid stains subjected to contamination and environmental insults," *J. Forensic Sci.* **36,** 1284–1298.

Von Beroldingen, C. H., Blake, E. T., Higuchi, R., Sensabaugh, G., and Ehrlich, H. (1989). "Application of PCR to the analysis of biological evidence. *In* "PCR Technology: Principles and Applications for DNA Amplification" (H. A. Ehrlich, ed.), Stockton Press, New York, pp. 209–223.

Blake, E., Mihalovich, J., Higuchi, R., Walsh, P. S., and Ehrlich, H. (1992). "Polymerase chain reaction (PCR) amplification of human leukocyte antigen (HLA)-DQ oligonucleotide typing on biological evidence samples: casework experience," *J. Forensic Sci.* **37,** 700–726.

Budowle, B., Moretti, T. R., Baumstark, A. L., Defebaugh, D. A., and Keys, K. (1999). "Population data on the thirteen CODIS core short tandem repeat loci in African Americans, US Caucasians, Hispancics, Bahamians, Jamaicans and Trinidadians," *J. Forensic Sci.* **44,** 1277–1286.

Dieffenbach, C. W., and Dveksler, G. S. (1993). "Setting up a PCR laboratory," *PCR Methods Appl.* **3,** 2–7.

Frégeau, C. J., and Fourney, R. M. (1993). "DNA typing with fluorescently tagged short tandem Repeats: A sensitive and accurate approach to human identification," *BioTechniques* **15,** 100–119.

Haugland, R. P. (1996). "Handbook of Fluorescent Probes and Research Chemicals, 6th Edition," Molecular Probes, Eugene, OR, pp. 144–156.

Jeffreys, A., Wilson, V., and Thein, S. L. (1985). "Hypervariable 'minisatellite' regions in human DNA," *Nature* **314,** 67–73.

Kimpton, K. P., Gill, P., Walton, A., Urquhart, A., Millinan, E. A., and Adams, M. (1993). "Automated DNA profiling employing multiplex amplification of short tandem repeat loci," *PCR Methods Appl.* **3,** 13–22.

Laber, T. L., O'Connor, J. M., Iverson, J. T., Liberty, J. A., and Bergman, D. L. (1992). "Evaluation of four Deoxyribonucleic Acid (DNA) extraction protocols for DNA yield and variation in restriction fragment length polymorphism (RFLP) sizes under varying gel conditions," *J. Forensic Sci.* **37,** 404–424.

Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982). "Molecular Cloning: A Laboratory Manual," Cold Spring Harbor Laboratory, NY, pp. 458–459, 469.

McNally, L., Shaler, R. C., Baird, M., Balazs, I., Kobilinsky, L., and De Forest, P. (1989). "The effects of environment and substrata on deoxyribonucleic acid (DNA): The use of casework samples from New York city," *J. Forensic Sci.* **34,** 1070–1077.

Miller, S. A., Dykes, D. D., and Polesky, H. F. (1988). "A simple salting out procedure for extracting DNA from human nucleated cells," *Nucl. Acids Res.* **16,** 1215.

Mullis, K. B., Faloona, F. A., Scharf, S. J., Saiki, R. K., Horn, G. T., and Erlich, H. A. (1986). "Specific enzymatic amplification of DNA *in vitro*: The polymerase chain reaction," *Cold Spring Harbor Symp. Quant. Biol.* **51,** 263–273.

Shows, T. B., McAlpine, P. J., and Bouchix, C., *et al.* (1987). "Guidelines for human gene nomenclature. An international system for human gene nomenclature (ISGN,1987)," *Cytogenet. Cell Genet.* **46,** 11–28.

Walsh, P. S., Metzger, D. A., and Higuchi, R. (1991). "Chelex® 100 as a medium for simple extraction of DNA for PCR based typing from forensic material," *BioTechniques* **10,** 506–513.

Walsh, P. S., Varlaro, J., and Reynolds, R. (1992). "A rapid chemiluminescent method for quantitation of human DNA," *Nucl. Acids Res.* **20,** 5061–5065.

Waye, J. S., and Willard, H. F. (1986). "Structure, organization and sequence of alpha satellite DNA from human chromosome 17: evidence for evolution by unequal crossing-over and an ancestral pentamer repeat shared with the human X chromosome," *Mol. Cell. Biol*. **6,** 3156–3165.

Wyman, A., and White, R. (1980). "A highly polymorphic locus in human DNA," *Proc. Nat. Acad. Sci. (USA)* **77,** 6754–6758.

Gene Expression, Regulation of

Göran Akusjärvi
Uppsala University

I. Introduction
II. Definition of a Transcription Unit
III. Regulation of Transcription in Eukaryotes
IV. Regulation of Transcription in Prokaryotes
V. Posttranscriptional Regulation of Gene Expression

GLOSSARY

Exon In eukaryotes, the part of the precursor-RNA that reaches the cytoplasm as part of a mRNA, rRNA, or tRNA (see also intron).

Intron The part of the precursor-RNA that is removed during RNA splicing, before the mature mRNA, rRNA, or tRNA is transported to the cytoplasm (see also exon).

Nucleosome The basic structural unit used to condense DNA in a cell. The nucleosome consists of a disc-shaped core of histone proteins (H2A, H2B, H3, and H4) around which an approximately 146-base-pair segment of DNA is wrapped.

Precursor-RNA The nuclear RNA transcript produced by transcription of the DNA. The precursor-RNA contains both exonic and intronic sequences. Introns are removed by RNA splicing before the mature RNA is transported to the cytoplasm.

Promoter The DNA sequence element that determines the site for transcription inititation for an RNA polymerase.

RNA splicing The nuclear process by which introns are removed from the precursor-RNA before the mature mRNA, rRNA, or tRNA is transported to the cytoplasm.

U snRNP An RNA protein complex consisting of uridine-rich small nuclear RNAs (U snRNAs) complexed with a common set of proteins, the Sm proteins, and U snRNA-specific proteins. The U1, U2, U4, U5, and U6 snRNPs are involved in RNA splicing. Other U snRNPs serve other functions in the cell.

OUR GENES are stored in a stable DNA molecule that is faithfully replicated and transmitted to new daughter cells. Each cell in our body contains the same genetic information. The development of complex multicellular organisms such as humans with highly specialized organs and cell types then arises through a complex regulation of expression of the genetic material. Recent advances in whole-genome sequencing have suggested that the difference between a simple organism like a bacteria and a much more complex organism like a human may only result from as little as a 5- to 10-fold difference in the number of genes. Thus, a prototypical bacterial genome encodes for 2000–4000 genes, whereas the recently completed sequencing of

the human genome suggests a total gene number slightly more than 20,000. Although this estimate may be too low, it appears unlikely, based on other measurements, that the number of genes in humans will exceed 50,000. At a first, and even a second, glance this small difference in the number of genes makes it difficult to understand why humans and bacteria are so different from each other.

The past decade has seen an explosive increase in information about regulation of gene expression. This review summarizes some of the general themes that have emerged. It is focused on expression of protein-encoding genes in higher eukaryotes. At appropriate places a comparison with gene expression in prokaryotes is made in an attempt to highlight similarities and to show differences that might provide some answers to how mechanistic differences in the regulation of genes may provide at least part of the solution of to how a complex organism like humans may have arisen without an enormous increase in the number of genes compared to prokaryotes.

I. INTRODUCTION

Expression of the genetic information has been summarized in the so-called central dogma, which postulates that the genetic information in a cell is transmitted from the DNA to an RNA intermediate to protein. A major difference between simple and complex organisms is the existence of a cell nucleus. Thus, prokaryotes, which include the bacteria and the blue-green algae, do not have a nucleus, whereas eukaryotes, which include animals, plants, and fungi, have cells with a nucleus that encapsulates the DNA. The basic mechanisms to regulate gene expression in eukaryotes and prokaryotes are very similar, although eukaryotes generally use more sophisticated methods to squeeze out more information from the DNA sequence. In prokaryotes on–off switches of transcription appear to be the key mechanism to control gene activity, although other mechanisms also contribute to the control of gene expression: transcriptional attenuation, transcriptional terminations, and posttranscriptional effects. In eukaryotes similar mechanisms are in operation. However, a key difference between prokaryotes and eukaryotes is the extensive use of RNA processing to generate a mature mRNA. Thus, eukaryotic genes are encoded by discontinuous DNA segments that require a posttranscriptional maturation to produce a functional mRNA. As will be discussed later in this review, the requirement for RNA splicing may be a key to the development of a highly differentiated organism like humans. The general postulate that one gene makes one protein was derived from genetic studies of bacteriophages and does not apply to higher eukaryotes. Because of alternative RNA processing events a large fraction of eukaryotic genes encode for multiple proteins (see Section V).

II. DEFINITION OF A TRANSCRIPTION UNIT

A transcription unit represents the combination of regulatory and coding DNA sequences that together make up an expressible unit, whose expression leads to synthesis of a gene product that often is a protein but also may be an RNA molecule. In prokaryotes, proteins in a specific metabolic pathway are often encoded by genes that are clustered and transcribed into one polycistronic mRNA. A polycistronic mRNA encodes for multiple proteins. In such mRNAs, ribosomes are recruited to internal translational initiation sites through an interaction between the 16S ribosomal RNA and the so-called Shine–Delgarno sequence located immediately upstream of the translational start codon that is used to initiate protein synthesis.

In eukaryotes, in contrast, the primary transcription product is a precursor-RNA that undergoes several posttranscriptional maturation steps before it is transported to the cytoplasm and presented to the ribosomes. Thus, the 5′ end of the pre-mRNA is capped early after transcription initiation by addition of an inverted methylated guanosine nucleotide (the m7G-cap), the pre-mRNA is cleaved at its 3′ terminus, and an approximately 250-nucleotide poly(A) tail is added posttranscriptionally; finally, the pre-mRNA is spliced to remove the intervening intron sequences, and thus form the spliced mRNA which is transported to the cytoplasm. These posttranscriptional processing events give eukaryotes a unique, very important level to control gene expression (see Section V.). Furthermore, a eukaryotic mRNA usually is functionally monocistronic. This means that even if the mRNA encode for multiple open translational reading frames, the open reading frame closest to the 5′ end of the mRNA is typically the only one translated into protein. This results from the fact that the eukaryotic ribosome recognizes the mRNA by binding to the modified 5′ end of a mRNA (recognizing the cap nucleotide), whereas prokaryotic ribosomes recognizes internal Shine–Delgarno sequences in the polycistronic mRNA.

Transcription involves synthesis of an RNA chain that is identical in sequence to one of the two complementary DNA strands. DNA sequence elements upstream of the initiation site for transcription make up the promoter that binds the RNA polymerase responsible for synthesis of the precursor-RNA. Transcription can be subdivided into at least three stages: (1) initiation, which begins by RNA polymerase binding to the double-stranded DNA molecule and incorporation of the first nucleotide(s); (2)

elongation, during which the RNA polymerase, by a processive mechanism, moves along the DNA template in a 5′-to-3′ direction and extends the growing RNA chain by copying one nucleotide at a time; and (3) termination, where RNA synthesis ends and the RNA polymerase complex disassembles from the transcription unit.

Because of space limitation this review will mostly cover RNA synthesis and maturation of protein-encoding messenger RNAs (mRNAs). However, similar mechanisms are used to regulate synthesis of other types of RNA molecules.

III. REGULATION OF TRANSCRIPTION IN EUKARYOTES

Eukaryotic cells contain three DNA-dependent RNA polymerases which are responsible for synthesis of specific RNA molecules in the cell. RNA polymerase I is responsible for the synthesis of ribosomal RNA, RNA polymerase II is responsible for the synthesis of protein-coding mRNA and four small stable RNAs involved in splicing (U1, U2, U4, and U5), and RNA polymerase III is responsible for the synthesis of small RNAs such as transfer RNA (tRNA) and 5S RNA and a whole array of small RNAs including U6, which is involved in splicing. The three polymerases are large enzymes consisting of approximately 15 subunits each. Many subunits are shared among the different polymerases, whereas others are unique and determine the promoter specificity during the transcription process. All three eukaryotic RNA polymerases contain core subunits that show a great homology with the *Escherichia coli* RNA polymerase, suggesting that the basic mechanism of RNA synthesis evolved early during evolution and is conserved.

As will become important later in this review, the largest subunit of RNA polymerase II contains a 52-times-repeated stretch of seven amino acids at the carboxy-terminus. This heptapeptide repeat is refereed to as the carboxy-terminal domain (CTD) and contains serines and a tyrosine that contribute to transcriptional regulation via reversible phosphorylation. The RNA polymerase that assembles at the promoter contains an unphosphorylated CTD tail. This CTD tail anchors the polymerase to the promoter by making interactions with TFIID bound at the TATA box. The release of the RNA polymerase from the promoter (i.e., start of elongation) is associated with a phosphorylation of the CTD tail.

A. Structure of a Eukaryotic Promoter

The transcriptional activity of a prototypical RNA polymerase II gene is regulated by a series of DNA sequence elements that can be subdivided into the core promoter element, which consists of the transcriptional start site and the TATA element, and upstream regulatory elements, which are needed for regulated transcription (Fig. 1).

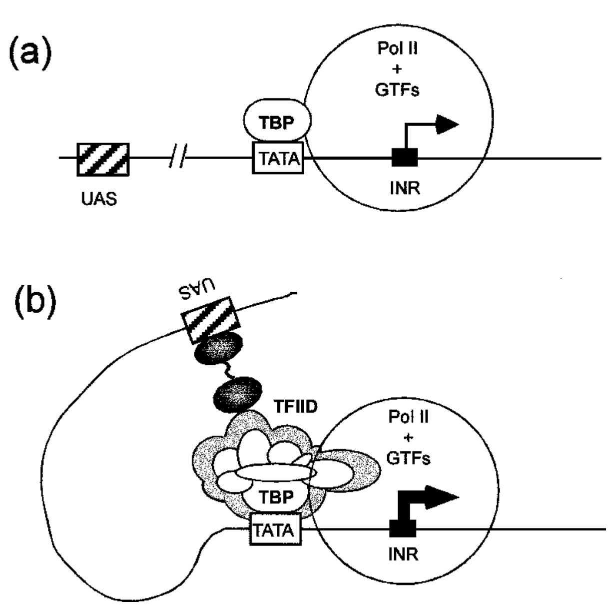

FIGURE 1 A schematic model for preinitiation complex formation on a core promoter (a) and a promoter regulated by enhancer-binding transcription factors (b). The figure is meant to illustrate that the TATA-binding factor TBP is sufficient for basal transcription, whereas enhancer-dependent transcription requires TFIID, which consists of TBP plus the TBP associated factors (TAFs). Pol II, RNA polymerase II; GTFs, general transcription factors; INR, initiator region.

1. The Core Promoter

The TATA element located 25–30 base pairs (bp) upstream of the transcription initiation site is critical for formation of the preinitiation complex by functioning as the binding site for the TATA-binding protein (TBP). The core promoter is sufficient to direct basal (unregulated) transcription. In some genes the transcriptional start site includes an initiator (Inr) element that binds specific factors that may substitute for the TATA box in recruiting the basal transcriptional machinery to the promoter. Although the core promoter is of fundamental importance for binding of the general transcription apparatus, the composition of elements may influence regulation of promoter activity.

2. Upstream Regulatory Factors

Upstream activating sequences (UAS), or transcriptional enhancer elements, are binding sites for transcription factors that stimulate RNA synthesis. The term UAS is used to describe DNA sequence elements that are located close to

the core promoter, so-called promoter proximal elements. UAS elements are typically located within 200 base pairs upstream of the transcription initiation site. Enhancer sequences are DNA segments containing binding sites for multiple transcription factors that activate transcription independent of their orientation and at a great distance [up to 85 kilobases (kb)] from the start site of transcription. Enhancer elements can be located either upstream or downstream of the transcription initiation site. Enhancer sequences activate transcription in a position-independent manner because they become spatially positioned close to the core promoter through bending of the DNA molecule (Fig. 1).

In addition to enhancer elements, eukaryotic promoters contain upstream repressor elements, which block RNA synthesis by various mechanisms by recruiting factors that interefere with enhancer factors or directly block RNA polymerase II recruitment. A third class of DNA sequence elements regulating transcription are the transcriptional silencers. A classical silencer represses transcription in a position- and orientation-independent fashion. The silencer element is thought to block transcription by functioning as the nucleation site for binding of histones or silencing proteins that coat the region, thereby making the promoter inaccessible for RNA polymerase recruitment.

The human genome encodes for several thousand different transcription factors. Promoters that contain combinations of binding sites for different transcription factors regulate different genes. Thus, for example, a gene specifically expressed in the liver or the brain uses liver- or brain-specific enhancer binding transcription factors, respectively, to achieve a tissue-specific gene expression. The basal transcriptional machinery appears to a large extent to be the same in all cell types.

B. Regulation of Promoter Activity

From a regulatory point of view it is important to note that TBP is sufficient to recruit RNA polymerase II and direct basal transcription from the core promoter. However, the basal transcription factor TFIID has been shown to play a central role in activated transcription by binding to the TATA element in the core promoter and facilitating the recruitment of the RNA polymerase holoenzyme to the promoter (Fig. 1). TFIID is a multiprotein complex consisting of TBP and approximately 11 TBP-associated factors (TAFs). The TAFs have been shown to be essential for regulated transcription by mediating contact with enhancer binding factors. Thus, TBP is sufficient for constitutive transcription but TAFs are necessary for regulated transcription (Fig. 1). *In vitro* studies suggest that assembly of an initiation-competent RNA polymerase at a promoter can be subdivided into several steps where different basal transcription factors are sequentially recruited to the promoter. However, fractionation experiments have shown that on certain promoters the RNA polymerase and most or all of the general transcription factors may be recruited as a single complex. *In vivo* activation of the thousands of promoters present in the human genome may use a large spectrum of mechanistic possibilities.

An important finding was the observation that UAS-binding transcription factors are modular in structure with a DNA-binding domain and an effector domain that could be exchanged without losing their predicted biological activity. The effector or activation domains in different transcription factors perform the same task but have different properties, for example, consisting of acidic blobs, proline-rich, glutamine-rich, or serine/threonine-rich sequences. Different classes of UAS-binding transcription factors may transmit a signal to the basal promoter complex by making specific contacts with different TAFs. For example, an interaction between the UAS-binding transcription factor SP1 and TAF-110 has been shown to be necessary for SP1-mediated activation of transcription. Collectively stabilized protein–protein interactions between UAS-binding factors and the general transcriptional factor TFIID are likely to facilitate recruitment of the RNA polymerase to the core promoter element, and as a consequence increase the transcriptional activity of the promoter (Fig. 1).

Transcription factors can be subdivided into families based on the structural feature of the DNA-binding domain. Thus, the DNA-binding domain may interact with the DNA through structural types like the helix-turn-helix motif found in homeodomain proteins, zinc fingers, or leucine-zipper-basic DNA-binding domain motifs. Heterodimerization between members of UAS-activating transcription factors belonging to such structural types is not uncommon and has been shown to increase the repertoire by which transcription factors can interact with different promoter sequences. For example, the prototypical AP1 transcription factor, which belong to the leucine-zipper family of transcription factors, consists of a heterodimer of c-*jun* and c-*fos*. It binds to its cognate DNA motif with a higher affinity than, for example, a c-*jun*–c-*jun* homodimer or a *JunB*–c-*fos* heterodimer. The combinatorial complexity is further increased by the fact that c-*jun* may form heterodimers with members of the ATF family of transcription factors. Thus, heterodimerization between different members of a transcription factor family is an important mechanism to generate factors with alternative DNA-binding specificity.

When the RNA polymerase leaves the promoter, TFIID remains bound at the TATA element and is ready to help a second RNA polymerase to bind and initiate transcription at the same promoter. The activity of TFIID appears also

to be regulated by inhibitory proteins that interact with TBP. Such TBP–inhibitory protein complexes may serve an important regulatory role by keeping genes which have been removed from inactive chromatin in a repressed but rapidly inducible state.

1. Regulation of Transcription by Chromatin Remodeling

During the last decade a wealth of information has demonstrated the significance of the chromatin template for the transcriptional activity of a promoter. It has been known for decades that the DNA in our cells is wrapped around a protein core called the nucleosome. The nucleosome consists of two copies each of four histones: H2A, H2B, H3, and H4. The DNA is packaged into either a loose structure called euchromatin or a more highly ordered structure called heterochromatin. The heterochromatin fraction is transcriptionally inactive, whereas active genes are found in the euchromatin fraction of the DNA.

Histones are modified by acetylation, phosphorylation, methylation, and ubiqutination. During recent years an impressive amount of work has demonstrated the significance of reversible histone acetylation as a regulatory mechanism controlling gene expression. Several lysines on the amino-terminal tail of each core histone can be acetylated. Lysines are negatively charged and make strong interaction with the phosphate backbone of DNA, thereby preventing basal transcription factors like TBP from interacting with DNA. Acetylation of lysines neutralizes this negative charge and reduces the electrostatic interaction of the histones with the DNA, thereby making the promoter region accessible for interaction with the basal transcription machinery (Fig. 2). A considerable amount of work shows that a general theme in transcriptional regulation is that acetylation of core histones results in looser nucleosomal structure, which makes the DNA more accessible for binding of transcription factors, and hence a gene more transcriptionally active. In contrast, histone deacetylation has the opposite effect and functions as a signal to repress transcription (Fig. 2).

Several transcriptional enhancer proteins have been shown to activate transcription by binding so-called coactivator proteins which have histone acyltransferase (HAT) activity. The best characterized are Gcn5 (yeast), $TAF_{II}250$, CBP, and p300. $TAF_{II}250$, which is a component of the TATA-binding basal transcription factor TFIID, may activate transcription by inducing acetylation of histones located in the vicinity of the TATA box. On the other hand, transcriptional repressor proteins have often been shown to inhibit RNA synthesis by recruiting histone deacetyltransfereases (HDACs), which cause a condensation of nucleosomes to a more compact, transcriptionally

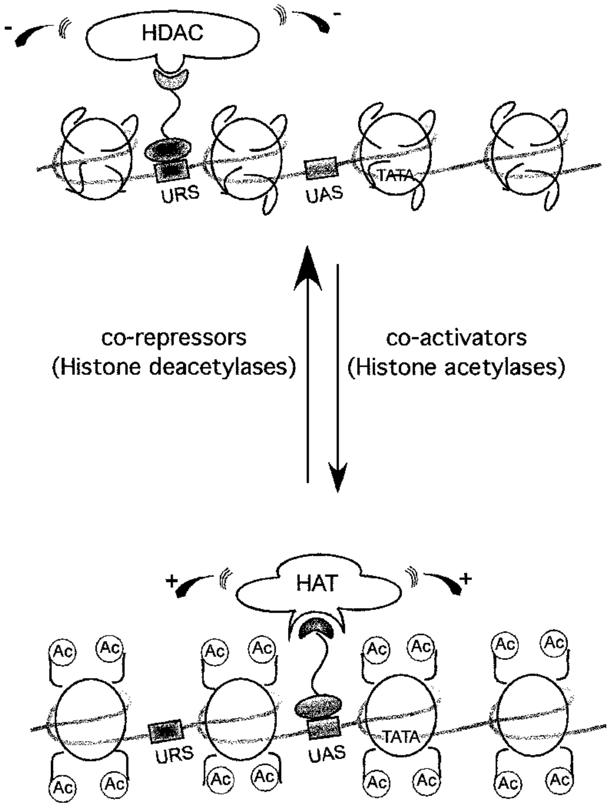

FIGURE 2 Role of the nucleosome in gene expression. Recruitment of histone deacetylases (HDACs) to a promoter inhibits binding of general transcription factors to the TATA element, thereby blocking transcription. Recruitment of histone acetylases (HATs) to the promoter results in acetylation of the amino-terminal tails of the core histones, thereby facilitating binding of the general transcription factors required for initiation of transcription. URS, upstream repressor sequence; UAS, upstream activating sequence.

inactive structure (Fig. 2). In yeast, HATs and HDACs are found in multiprotein complexes such as SAGA and Sin3 complexes, respectively. The equivalent, and additional, multienzyme complexes are also found in higher eukaryotes.

In addition, the nucleus contains so-called chromatin remodeling factors, such as the Swi/Snf complex, which has the capacity to reposition nucleosomes and transiently dissociate the DNA from the surface of the nucleosome. Depending on the promoter context, chromatin remodeling factors may cause an activation or repression of transcription.

It is likely that other histone modifications, such as phosphorylation, ubiquitinilation, and methylation, also play a significant regulatory role in transcriptional control of promoter activity, although the importance of these modifications has not yet been characterized to the same extent as has that of reversible acetylation. The main conclusion from these studies is that a linear assessment of the

DNA sequence elements capable of binding transcriptional activator or repressor proteins only tells us part of the story, namely which factors have the capacity to control promoter activity. However, actual RNA synthesis requires a complex interplay between UAS-binding factors and the chromatin or the chromatin remodeling factors that have positive or negative effects on promoter activity.

2. Regulation of Transcription Factor Activity

The activity of a UAS-binding transcription factor is subjected to a posttranslational regulation. There are in principle three ways that the activity of an UAS-binding transcription factor may be tuned (Fig. 3); covalent (like phosphorylation) or noncovalent (like hormone binding) modification of the UAS-binding factor, or variation of the subunit composition (like binding of an inhibitory protein). These mechanisms may be used individually or in combination with other mechanisms to regulate transcription. To illustrate the flexibility of transcriptional control in eukaryotic cells two examples are presented. The first example concerns the activation of steroid hormone-dependent gene transcription (Fig. 4). Steroid hormones are a group of substances derived from cholesterol which exert a wide range of effects on processes such as growth, metabolism, and sexual differentiation. A prototypical member of a steroid hormone-inducible transcription factor is the glucocorticoid receptor (GR). In the absence of hormone this receptor is found as a monomer in the cytoplasm complexed to the heat-inducible hsp90 protein. Treatment with steroid hormones results in the release of GR from hsp90, which renders GR free to dimerize and move to the nucleus, where it binds to its cognate DNA sequence element and activates transcription (Fig. 4). In addition to inducing a dissociation of the receptor from hsp90, ligand binding also induces a conformational change in the activation domain of GR, such that the activation domain binds transcriptional coactivator proteins that stimulate transcription. Interestingly,

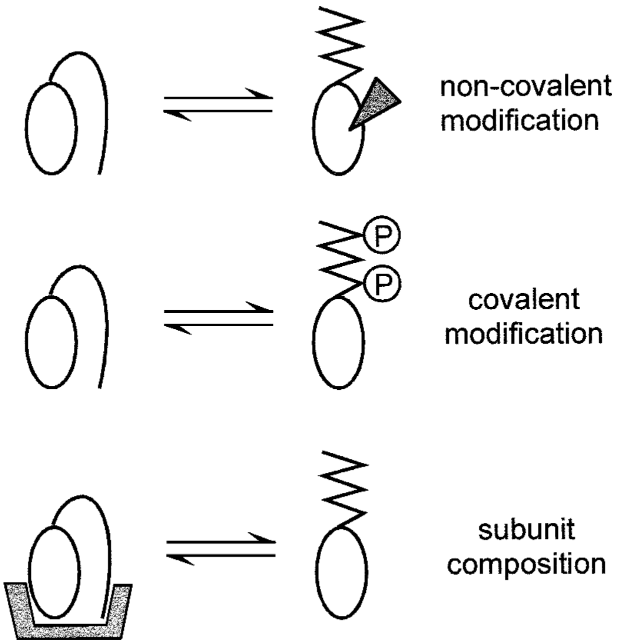

FIGURE 3 Three common mechanisms to regulate transcription factor activity.

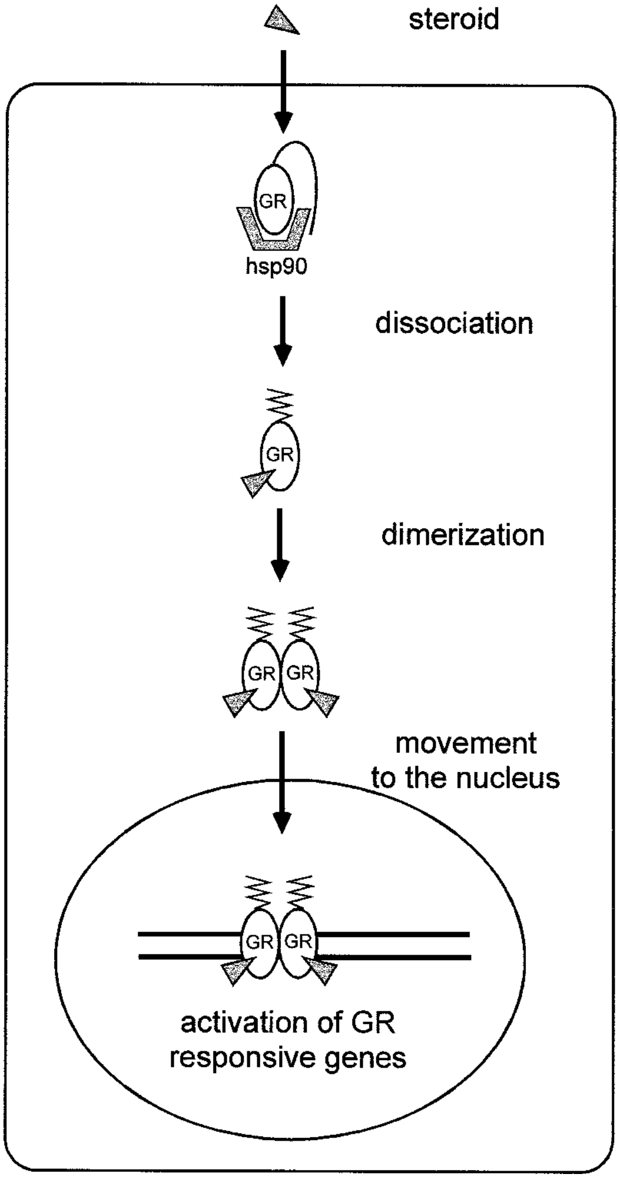

FIGURE 4 A schematic drawing showing the activation of glucocorticoid receptor (GR) by steroid hormone binding, which results in the dissociation of the cytoplasmic GR–hsp90 complex, followed by GR dimerization and translocation of GR to the nucleus.

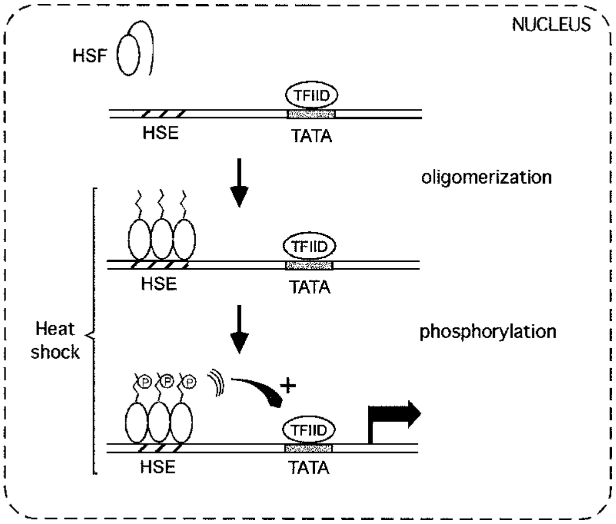

FIGURE 5 Activation of transcription of heat-shock genes in mammalian cells. The heat shock factor (HSF) is present as a monomer in normal cells. An increase in temperature results in a trimerization of HSF, which binds to the heat shock element (HSE). HSF is activated as a transcriptional enhancer protein by phosphorylation.

heat-shock proteins do not regulate other nuclear hormone receptors, such as the retinoic acid and thyroid hormone receptors. Thus, these receptors bind DNA in the absence of the ligand. In this case ligand binding results in a conformational change of the activation domain permitting binding of coactivator proteins.

The second example concerns heat-shock activation of transcription in mammalian cells (Fig. 5). When cells are subjected to an elevated temperature (heat shock) they respond by activating synthesis of a small number of genes encoding for so-called heat-shock proteins. These proteins serve an important function during heat shock by binding to cellular proteins, which become denatured by the increase in temperature. Subsequently, the heat-shock proteins help to renature the proteins to their native conformation. Transcription of heat-shock genes is controlled by the heat-shock transcription factor (HSF), which binds to the heat-shock element (HSE) found in the promoter of all genes regulated by heat shock. HSF is activated by two mechanisms (Fig. 5). Thus, in normal cells HSF exists as a monomer. An increase in temperature results in unfolding of HSF, which exposes the DNA-binding domain and allows it to bind to other HSFs and form a trimer that binds to the HSE. However, binding of HSF to DNA is not enough to activate transcription. Thus, HSF needs to be modified by phosphorylation before it activates transcription of the heat-shock genes. Interestingly, TFIID is bound to the TATA element in heat-shock genes also in uninduced cells. Thus, binding of an active HSF to the HSE is needed for recruitment of the RNA polymerase to the heat-shock promoter. The binding of TFIID to the uninduced promoter may help heat-shock genes respond more rapidly to an increase in temperature.

Cell type and differentiation-specific gene expression is often regulated by the availability of specific transcription factors. Genes that are expressed in specific organs contain binding sites for cell type-specific transcription factors. Thus, tissue-specific transcription is often regulated by the precise arrangement of regulatory UAS motifs in the promoter, the availability of the cognate transcription factors, and the way these transcription factors influence the activity of the promoter. Thus, for example, the liver and the brain encode for respectively liver- and brain-specific transcription factors that ensure a tissue-specific expression gene expression. Since transcription factors are typically dimeric proteins, the exact composition of the two partners may vary among cell types and have different transcription regulatory properties.

3. Regulation of Transcription Elongation

Although transcription initiation has only been discussed, RNA polymerase elongation is also an important step in regulating gene expression in eukaryotes. Thus, there are several examples where the RNA polymerase halts at specific pause sites during elongation. To be able to complete the synthesis of the precursor-RNA the polymerase has to be able to override this attenuation of transcription. The best-characterized example is the human immunodeficiency virus (HIV) Tat protein, which binds to a stem-loop structure at the 5′ end of the HIV transcript, the TAR sequence. In the absence of the Tat protein, HIV transcription terminates approximaqtely 50 nucleotides downstream of the initiation site. When Tat is present it binds to the TAR sequence and recruits a cyclin T/Cdk9 complex which is responsible for phosphorylation of the CTD tail of RNA polymerase II, thereby alleviating termination and permitting the RNA polymerase to synthesize the full-length HIV genomic RNA.

IV. REGULATION OF TRANSCRIPTION IN PROKARYOTES

A. Introduction

The mechanisms to initiate transcription in eukaryotes and prokaryotes are similar. As a comparison to control of transcription in eukaryotes some key features in transcriptional control in bacteria will be given. Prokaryotic cells contain only one type of RNA polymerase, which is responsible for synthesis of all types of RNA: mRNA, rRNA,

and tRNA. The core polymerase is a four-subunit enzyme consisting of two a, one b, and one b′ subunit. However, the holoenzyme, which is the complete enzyme, contains the core polymerase plus the sigma factor, which may regarded as the prokaryotic equivalent of the general transcription factors found in eukaryotes. The sigma factor is required for proper RNA polymerase binding to a prokaryotic promoter. After initiation of transcription the sigma factor leaves the polymerase complex and elongation is taken care of by the core polymerase.

Similar to a eukaryotic promoter a prototypical prokaryotic core promoter contains a conserved TATAAT located at position −10 relative to the transcription start site and resembling the eukaryotic TATA element. In addition, prokaryotic promoters contain a conserved TTGACA located at position −35. The spacing between the two elements is of critical importance for the efficiency by which the RNA polymerase binds to the promoter. The exact sequences at the −10 and −35 positions vary slightly for different transcription units. Usually promoters that have a better homology to the consensus sequences also initiate transcription more efficiently. An important mechanism to regulate the transcriptional activity of a prokaryotic promoter is to provide the core polymerase with different sigma factors. Thus, different sigma factors determine promoter specificity by recognizing −10 and −35 elements with different base sequences. This strategy mediates the heat shock response and the regulated expression of genes during developmental processes. For example, sporulation in *Bacillus subtilis* uses a cascade of different sigma factors to cause the transformation of a vegetative bacterium to a spore.

The existence of transcriptional enhancers similar to those found in eukaryotic cells has also been described in prokaryotes. For example, the enhancer-binding protein nitrogen regulatory protein C (NTRC) in the *glnA* promoter from *Salmonella typhimurium* activates transcription from a distance by means of DNA looping. NTRC stimulates transcription by a transient contact between the activator and the polymerase. This catalyzes an unwinding of the DNA at the promoter, which then allows the RNA polymerase to initiate transcription.

B. The *Lac* Operon

Transcriptional repression is a key mechanism to control the activity of prokaryotic promoters. Enzymes used in a specific metabolic pathway are often organized into an operon that is transcribed into a single polycistronic mRNA. Specific repressor proteins then control the transcriptional activity of the operon by regulating RNA polymerase binding to the promoter. Repressor proteins are DNA-binding proteins that typically block RNA polymerase access to the −10 and/or −35 regions in the promoter or transcription elongation by associating with an operator sequence that is positioned downstream of the start site of transcription. Usually these regulatory proteins undergo allosteric changes in response to binding of a specific ligand. The paradigm of a prokaryotic operon regulated by a specific repressor protein is the *lac* operon in *E. coli.* In this system synthesis of proteins necessary for usage of lactose as a carbon source is repressed by the *lac* repressor protein if cells have the possibility to use glucose for growth. Thus, in the presence of glucose the *lac* repressor binds to its operator sequence, which overlaps the transcription start site in the *lac* operon (Fig. 6), and blocks RNA polymerase binding to the *lac* promoter. If cells are grown on lactose as the carbon source, lactose functions as an inducer of *lac* operon transcription by binding to the *lac* repressor and converting it to an inactive form that does not bind DNA (Fig. 6) and therefore is unable to inhibit transcription of the *lac* operon. The polycistronic *lac* mRNA encodes for the specific proteins necessary for metabolism of lactose. The *lac* operon represents an example of an inducible system where an inducer activates transcription. However, inducers can also have the opposite effect and repress transcription of an operon, like the *trp* operon in *E. coli.*

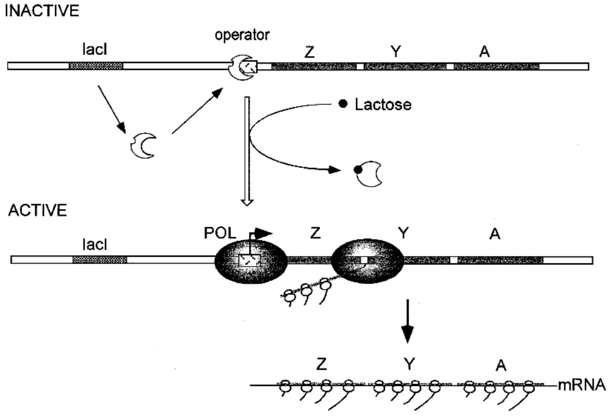

FIGURE 6 Regulation of the *lac* operon in *E. coli*. The *lac I* gene encodes for a transcriptional repressor protein that binds to an operator sequence in the *lac* operon, thereby preventing synthesis of the structural genes required for metabolism of lactose. If *E. coli* is grown on lactose as the sole carbon source, lactose binds to the *lac I* repressor protein and inactivates it as a repressor of *lac* operon transcription. As a consequence, the β-galactosidase (*lac Z*), the permease (*lac Y*), and the β-galactosidase transacetylase (*lac A*) enzymes are synthesized.

V. POSTTRANSCRIPTIONAL REGULATION OF GENE EXPRESSION

Expression of eukaryotic genes is not only controlled at the level of initiation of RNA synthesis. Thus, the

precursor-RNA synthesized by RNA polymerase II undergoes several posttranscriptional modifications before a mature mRNA is formed. For example, the transcript is capped at its 5′ end, the 3′ end is generated by a specific cleavage polyadenylation reaction, and intronic sequences are removed by RNA splicing.

Early after initiation of transcription, when the nascent RNA chain is 25–30 nucleotides long, the 5′ end is modified by addition of an inverted 7-methylguanosine, the cap nucleotide. The capping enzymes are brought to the transcribing polymerase by specific association with the hyperphosphorylated form of the CTD tail on RNA polymerase II. As mentioned above, the CTD tail becomes phosphorylated when RNA polymerase II progress from the initiation to the elongation phase of RNA synthesis. Since RNA polymerases I and III do not have a CTD tail, only RNA polymerase II transcripts are capped. The cap plays a crucial role in initiation of translation by binding the translational initiation factor eIF4F required for the recruitment of the small subunit of the ribosome to the mRNA. The translational start site is then identified by a scanning mechanism where the ribosome usually selects the first AUG triplet as the start codon for protein synthesis. The selective addition of a cap to RNA polymerase II transcripts therefore provides a logical explanation to why this class of RNAs is used for translation. Polyadenylation and RNA splicing are key mechanisms to regulate eukaryotic gene expression and are therefore described in more detail below.

A. Exons and Introns: General Considerations

Virtually all prokaryotic genes are encoded by a collinear DNA sequence: the concept one gene, one mRNA. In contrast, most eukaryotic genes are discontinuous, with the coding sequences (exons) interrupted by stretches of noncoding sequences (introns). Introns are present at the DNA level and in the primary transcription product of the gene (the precursor-RNA), and are removed by RNA splicing before the mature mRNA is transported to the cytoplasm. Recent experiments suggest that splicing is necessary for efficient transport of intron-containing precursor-RNAs. Introns have been found in all types of eukaryotic RNA—mRNA, rRNA, and tRNA. Because of space limitation, only introns in protein-encoding genes will be described.

The number of introns in mRNA-encoding genes varies considerably among genes. For example, c-*jun*, histone, heat-shock, and the α-interferon genes have no introns, whereas the gene for dystrophin has more than 70 introns. Also, the size of introns can vary from less than 100 nucleotides to several million nucleotides in length. The extreme example is the *Drosophila Dhc*7 gene, which contains a 3.6 million-nucleotide-long intron. This intron is approximately double the size of most bacterial genomes and takes days to transcribe. In contrast, exons are typically short, usually less than 350 nucleotides. This comes from the fact that splice sites used to define the borders of the splicing reaction are defined across the exon, not the intron—the so-called exon definition model (see below). Some eukaryotic genes are remarkably large. For example, the human gene for dystrophin covers approximately 2.4 million base pairs. The RNA polymerase that initiates transcription requires approximately 20 hr to synthesize the full-length precursor-RNA. Subsequently, more than 99.5% of the transcript is removed by RNA splicing. Thus, the final mRNA that is transported to the cytoplasm is only around 14,000 nucleotides. The extreme lengths of eukaryotic genes place a high demand on the stability of the transcription complex. Thus, an RNA polymerase that binds to a promoter must stay attached for days with the DNA template to be able to complete synthesis of the longest genes.

It is interesting to note that introns in eukaryotic genes almost always interrupt the protein-coding portion of the precursor-RNA. Thus, introns are rarely found after the translational stop codon, within the 3′ noncoding portion of the mRNA. This organization is significant since the presence of an intron downstream of the translational stop codon in a reading frame is sensed as a signal that the precursor-RNA has been incorrectly spliced or for other reasons is defective, and will not produce the correct protein after translation in the cytoplasm. Such nuclear transcripts are sent for destruction by a mechanism that is collectively called the non-sense-mediated mRNA decay mechanism. How the translational reading frame is read already in the nucleus is not known. The easiest explanation would be that there exists a nuclear ribosome-like structure that scans the spliced mRNA for a full-length translational reading frame before the mRNA is transported to the cytoplasm. However, this question is controversial and has not been proven.

1. Mechanism of RNA Splice Site Choice During Spliceosome Assembly

The sequence elements used to specify the splice sites are remarkably short and degenerate in a eukaryotic precursor-RNA. Thus, short conserved sequence motifs at the beginning (5′ end) and the end (3′ end) of the intron guide the assembly of a large RNA protein particle, the spliceosome (Fig. 7), which catalyzes the cleavage and ligation reactions necessary to produce the mature cytoplasmic mRNA. The nucleus of eukaryotic cells contains several abundant low-molecular-weight RNAs, so-called U snRNAs. The U snRNAs derive their name from the fact that they were initially characterized as RNAs rich in uridines. Five of these U snRNAs (U1,

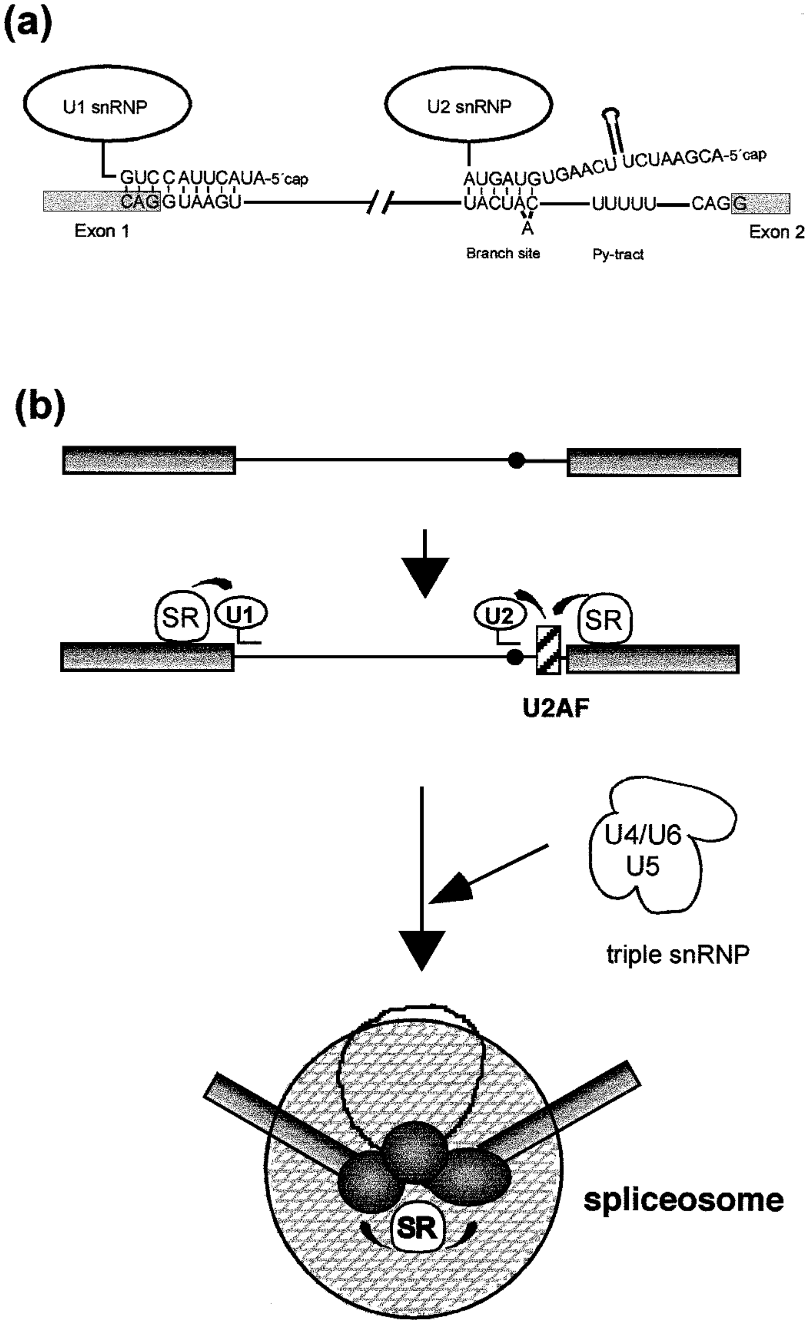

FIGURE 7 A simplistic model for spliceosome assembly. (a) The 5′ splice site and the branch site are defined via a direct base pairing between the RNA components of the U1 and U2 snRNPs and the precursor-RNA, respectively. (b) Efficient recruitment of the U snRNPs to the spliceosome is aided by non-snRNP proteins. Thus, SR proteins facilitate U1 snRNP binding to the 5′ splice site, whereas U2AF binds to the polypyrimidine tract at the 3′ splice site and helps U2 snRNP binding to the branch site. The U4/U6–U5 triple snRNP is recruited to form the mature spliceosome. SR proteins bring the 5′ and 3′ splice sites in close proximity for the catalytic steps of splicing by making simultaneous contact with splicing factors binding to the 5′ and 3′ splice sites, respectively (see also Fig. 8).

U2, U4, U5, and U6), ranging in size from 107 to 210 nucleotides, have been shown to participate in splicing. *In vivo* the snRNAs are found complexed to 6–10 proteins, generating the so-called small nuclear ribonucleoprotein particles (snRNPs). Some snRNP proteins are shared among different U snRNPs, whereas other snRNP proteins are unique to each U snRNP. During spliceosome assembly the ends of the introns are in part identified by RNA–RNA base pairing between the precursor-RNA and a U snRNP (Fig. 7). For example, the 5′ splice site is

recognized through a short base pairing between the U1 snRNA and the precursor-RNA. Similarly, a base pairing between U2 snRNA and the branch point defines the 3′ splice site. Later during spliceosome formation the U5–U4/U6 triple snRNP is recruited. In the triple snRNP, U4 and U6 snRNP form an extensive base pairing. The catalytically active spliceosome is generated by conformational changes, which results in a breakage of the base pairing between U4 and U6 snRNP and formation of new U–U snRNA and U snRNA–precursor-RNA base pairings. It is generally believed, although not proven, that the U snRNAs in the spliceosome are the enzymes that catalyze the two transesterification reactions required to excise the intron.

2. Non-snRNP Proteins Required for Splicing

The spliceosome, which is a large RNA–protein complex, with a size similar to a cytoplasmic ribosome, also contains numerous non-snRNP proteins which are important for correct splice site recognition. Assembly of the spliceosome proceeds over several stable intermediates (Fig. 7).

Efficient recruitment of U2 and U1 snRNP to the 3′ and 5′ splice sites also requires specific proteins. Here only two factors will be described. The first is U2 snRNP auxiliary factor (U2AF), which binds to the pyrimidine tract located between the branch site and the 3′ splice site in the precursor-RNA. U2AF stabilizes U2 snRNP binding to the branch site. The second factor is not one protein, but a family of proteins, designated SR proteins. SR proteins contain one or two amino-terminal RNA-binding domains and a carboxy-terminus rich in arginine (R) and serine (S) dipeptide repeats (the RS domain); hence the name SR proteins. Mechanistically, SR proteins appear to perform the same function in RNA splicing that transcriptional enhancer proteins do in transcription initiation. Thus, SR proteins bind to splicing enhancer sequences through their RNA-binding domains and stimulate spliceosome assembly by facilitating protein–protein interaction (Fig. 7). The RS domain functions as a protein interaction surface that makes contact with other SR proteins and so-called SR-related proteins. Thus, many proteins involved in RNA splicing contain RS domains. For example, SR proteins aid in efficient U1 snRNP binding to a 5′ splice site by interacting with the U1-70K protein, which is an RS-domain containing protein. However, in contrast to transcriptional enhancer proteins, which active transcription irrespective of the position where they bind, SR protein function is position dependent. Thus, in general, SR proteins function as splicing-enhancer proteins if they bind to the exon and function as splicing-repressor proteins if they bind to the intron in the precursor-RNA.

The number of SR proteins found in mammalian cells is surprisingly few considering the multitude of regulated splicing events for which they are required. Thus, only around 12 "true" SR proteins have been identified. Even more surprising, gene knockout experiments suggest that only one of the SR proteins is essential in *Caenorhabditis elegans*. Thus, disrupting the expression of the SR protein ASF/SF2 resulted in early embryonic lethality, whereas gene knockout of other SR proteins resulted in no change in phenotype. Probably, SR proteins show a large extent of functional redundancy, and disruption of one is compensated for by another SR protein. The essential role of SR proteins in spliceosome assembly makes them prime targets for regulation of gene expression.

3. The Exon Definition Model

The conserved sequences at the 5′ and 3′ ends of the intron are surprisingly short considering the precision by which very large introns are excised during splicing. The answer to this puzzle appears to be resolved by the fact that the 5′ and 3′ splice sites that are joined in the splicing reaction are not recognized over the intron. Instead splice sites are recognized across the exons—the so-called exon definition model. Thus, whereas introns can vary in length from less than 100 to more than 1 million nucleotides, internal exons in a precursor-RNA have a constant length and rarely exceed 350 nucleotides. The exon definition model postulates that U2 snRNP binding to a 3′ splice site makes contact with U1 snRNP binding to the downstream 5′ splice site (Fig. 8). If the 3′ and 5′ splice sites are too far away the model postulates that the intervening sequence is not recognized as an exon because U2 and U1 snRNP binding to respective splice sites cannot interact with each other. Once the exons have been defined in the precursor-RNA, adjacent exons are aligned for the splicing reaction.

B. Alternative RNA Splicing Is an Important Mechanism to Generate Protein Diversity

A major difference in gene regulation between a prokaryotic and a eukaryotic cell is the existence of mechanisms in eukaryotic cells that permit one gene to express multiple gene products. In bacteria a protein is encoded by a collinear DNA sequence. In contrast, in eukaryotes a single gene may encode for thousands of proteins. Thus, the discontinuous arrangement of eukaryotic genes, with introns interrupting the coding segments of the precursor-RNA, permit production of multiple, alternatively spliced mRNAs from a single gene. Examples of how a precursor-RNA can be alternatively spliced are shown in Fig. 9. This, of course, means that multiple proteins with different primary amino acid sequence and biological activity can be produced from a single eukaryotic gene. Of specific interest is that the production of alternatively spliced mRNAs in many cases is a regulated process, either in a temporal,

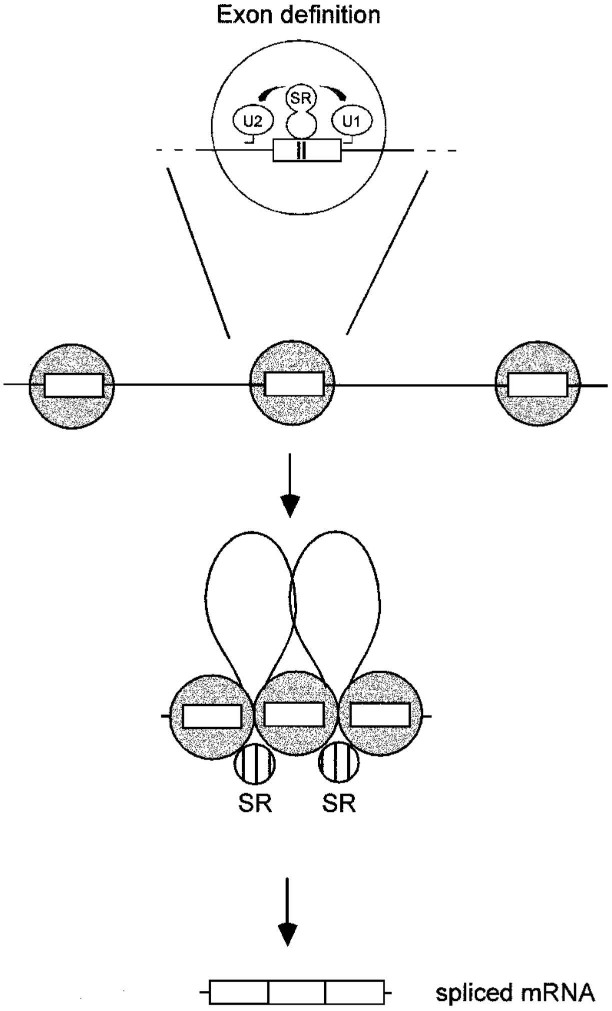

FIGURE 8 The exon definition model. Exons in a precursor-RNA are recognized as units by U2 snRNP (U2) binding to the 3′ splice site and U1 snRNP (U1) binding to the downstream 5′ splice site. Subsequently adjacent exons are defined across the intron. In both recognition steps SR proteins function as bridging proteins.

developmental, or tissue-specific manner. Changes in splicing have been shown to determine the ligand-binding specificity of growth factor receptors and cell adhesion molecules and to alter the activation domains of transcription factors. For example, the fibronectin precursor-RNA is alternatively spliced in hepatocytes and fibroblasts. In fibroblasts two exons which are skipped in hepatocytes are included during the splicing reaction. These two exons encode for protein domains that make fibroblast fibronectin adhere to many cell surface receptors. Fibronectin produced in hepatocytes lacks these two exons and therefore is translated to a hepatocyte-specific fibronectin protein that does not adhere to cells, allowing it to circulate in the serum.

The impact of alternative splicing on the coding capacity of a eukaryotic gene is mind-boggling. For example,

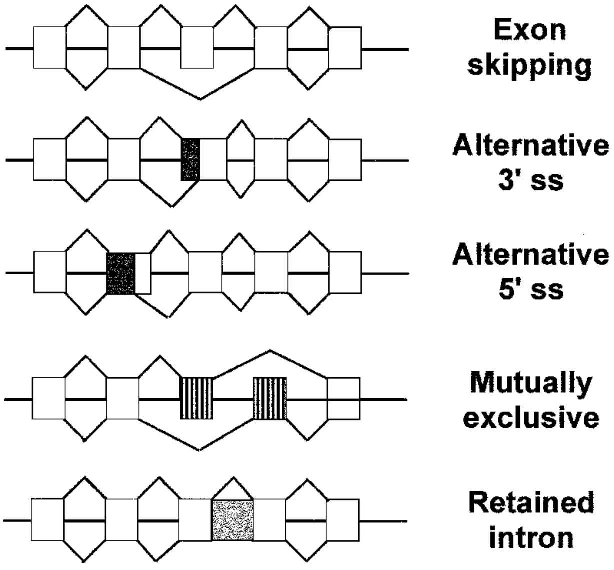

FIGURE 9 Examples of different patterns of alternative RNA splicing.

the *Drosophila DSCAM* gene, which encodes for an axon guidance receptor, has been estimated to produce 38,016 DSCAM protein isoforms by alternative splicing. This figure is remarkable since the total gene number calculated from the *Drosophila* DNA sequence suggests a total of only approximately 14,000 genes. Thus, a single *Drosphoila* gene produces almost three times the number of proteins compared to the number of genes in *Drosophila*. The *DSCAM* gene is not unique. There are many examples of human genes, like those for neurexins, *n*-cadherins, and calcium-activated potassium channels, that are known to produce thousands of functionally divergent mRNAs. A low estimate suggests that approximately 35% of all human genes produce alternatively spliced mRNAs. Thus, the estimate of 20,000–50,000 genes in the human genome could easily produce several hundred thousand, or million, proteins. Such differences in numbers are comforting because they make it easier to explain how a complex organism like humans with highly differentiated organs have evolved without an enormous increase in the number of genes compared to bacteria.

1. Regulation of Alternative RNA Splicing by Changes in SR Protein Activity

With a few exceptions little is known about the mechanistic details of how production of alternatively spliced mRNAs is regulated. However, it appears clear that the SR family of splicing factors partake in many regulated splicing events. SR proteins are highly phosphorylated, primarily within the RS domain. Thus, reversible RS domain phosphorylation has been shown to regulate SR protein

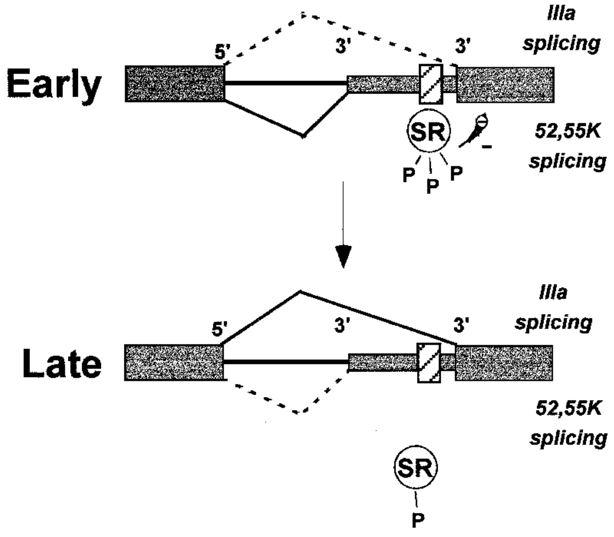

FIGURE 10 Regulation of alternative RNA splicing by SR protein phosphorylation. Hyperphosphorylated SR proteins present in normal cells and early virus-infected cell bind to a specific repressor element in the adenovirus L1 precursor-RNA and block spliceosome assembly at the IIIa 3′ splice site. This results in an exclusive production of the 52,55 K mRNA early after infection. Adenovirus induces a dephosphorylation of SR proteins late during infection, which alleviates the repressive effect of SR proteins on IIIa splicing, hence a shift to IIIa mRNA splicing.

interaction with other splicing factors and control alternative RNA splicing.

One of the best-characterized examples is the human adenovirus L1 unit (Fig. 10). The L1 unit produces two mRNAs, the 52,55K and the IIIa mRNAs, which are generated by alternative 3′ splice site selection. Splicing during an adenovirus infection is temporally regulated such that the IIIa mRNA is produced exclusively late during virus infection. It has been shown that highly phosphorylated SR proteins bind to an intronic repressor element and inhibit IIIa splicing during the early phase of infection. At late times of infection IIIa splicing is activated by a virus-induced dephosphorylation of SR proteins. This change in the phosphorylated status of SR proteins reduces their binding capacity to the repressor element and hence results in an alleviation of their repressive effect on IIIa 3′ splice site usage.

2. Maintenance of Sex in *Drosophila*: Sxl Regulation of Splicing

One of the most spectacular and best-characterized examples where alternative RNA splice site choice is used to regulate gene expression is the somatic sex-determination pathway in *Drosophila melanogaster* (Fig. 11). In this system sex determination has been shown to involve a cascade of regulatory events taking place at the level of alternative RNA splice site choice. The X chromosome encodes for transcription factors that control *Sex-lethal* (*Sxl*) transcription. In females, which contain two X chromosomes, the double dose of these transcription factors results in an activation of an early promoter of the *Sxl* gene. This promoter is inactive in males, which contain one X chromosome. The female-specific Sxl protein is an RNA-binding protein that binds to certain pyrimidine tracts and outcompetes U2AF binding to that site. Since the Sxl protein lacks the splicing activator function of U2AF, Sxl binding to a pyrimidine tract prohibits spliceosome formation at the 3′ splice site. Thus, once made, the female-specific Sxl protein autoregulates its own expression by ensuring that exon 3 in the Sxl pre-mRNA is efficiently skipped, thereby establishing a female-specific splicing of the Sxl pre-mRNA (Fig. 11). In male flies exon 3 is incorporated during splicing, resulting in the translation of a functionally inactive Sxl protein. This results from the fact that the third exon in the Sxl pre-mRNA contains a translational stop codon that causes a premature termination of translation. In addition, Sxl controls the splicing of downstream targets in the sex determination pathway. Thus, Sxl regulates splicing of the *transformer* (*Tra*) precursor-RNA by repressing usage of the male-specific 3′ splice site. Subseqently the female-specific Tra protein complexes with the Tra2 protein and activates a female-specific 3′ splice site in the *double-sex* (*Dsx*) precursor-RNA (Fig. 11). In males where a biologically inactive Sxl protein is expressed, the *Sxl*, *Tra*, and *Dsx* precursor-RNAs are processed by a default-splicing pathway, resulting in the development of male flies. The Dsx protein, the final protein in the cascade, is a transcription factor. The male- and female-specific Dsx proteins regulate development of flies along the male- or female-specific pathways.

It is widely accepted that alternative splicing is an important mechanism to regulate gene expression during growth and development in eukaryotic cells. A large number of eukaryotic genes have been shown to mature alternatively spliced mRNAs, examples include growth factors, growth factor receptors, intracellular messengers, transcription factors, oncogenes, and muscle proteins. The number of examples is constantly increasing.

C. The Basic Mechanism of 3′ End Formation

1. Introduction

In eukaryotes all mRNAs except the histone mRNAs have a 200- to 250-long 3′ poly(A) tail. It is noteworthy that the 3′ end of a eukaryotic mRNA is not generated by termination of transcription. Thus, the RNA polymerase continues to synthesize RNA beyond the actual 3′ end of the mature mRNA. Sequence analysis of a number of RNA polymerase II genes has reveled two elements that

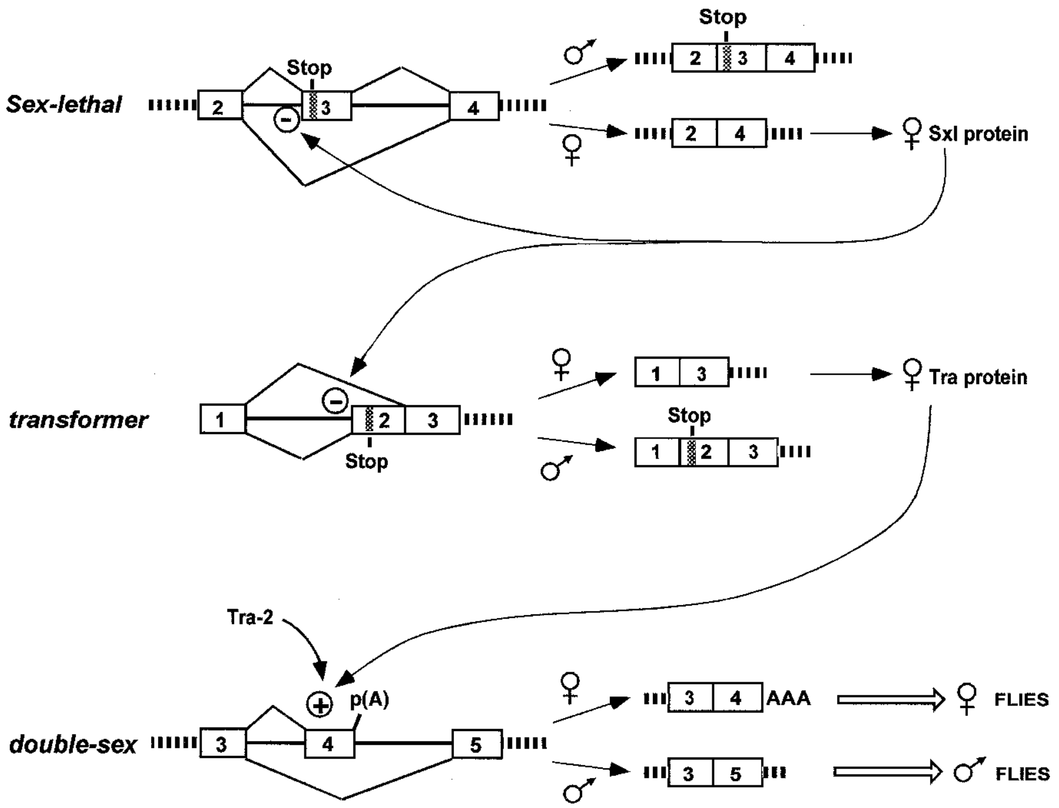

FIGURE 11 The cascade of regulated alternative splicing events controlling *Sxl, transformer,* and *double-sex* expression in *Drosophila melanogaster.* The positions of translational stop codons that will cause premature termination of protein synthesis are indicated by Stop. See text for further details.

by biochemical assays have been shown to specify the position of the 3′ end of a mRNA (Fig. 12). Thus, an almost invariable AAUAAA sequence located 25–30 nucleotides upstream of the cleavage site and a GU-rich sequence located within 50 nucleotides downstream of the cleavage–polyadenylation site are critical for 3′ end formation. The AAUAAA sequence, which resembles the AT-rich TATA box important for transcription initiation, binds the essential cleavage–polyadenylation specificity factor (CPSF). The GU-rich sequence binds the cleavage–stimulatory factor (CStF). In addition, two cleavage factors, CFI and CFII, and the poly(A) polymerase (PAP) assemble to form an active enzyme complex that cleaves the growing RNA chain and catalyzes the addition of the 200- to 250-nucleotide poly(A) tail. The transcribing RNA polymerase terminates RNA synthesis in an ill-defined sequence downstream of the cleavage–polyadenylation site. However, the cleavage–polyadenylation reaction has been shown to be required for transcription termination, suggesting that breakage of the primary transcript is coupled to termination of transcription, possibly through the action of 5′ exonucleases that degrade the downstream RNA chain generated by the cleavage reaction. Mechanistically, this may be analogous to Rho-dependent transcription termination in bacteria.

2. Regulation of Gene Expression at the Level of Alternative Poly(A) Site Usage

In addition to control of alternative RNA splicing, eukaryotic cells frequently use alternative poly(A) site usage to further increase the coding capacity of a gene. As described above, the 3′ end of a eukaryotic mRNA is generated by an endonucleolytic cleavage of the primary transcript rather than termination of transcription. This means that the RNA polymerase transcribing a gene may pass by several potential poly(A) sites before terminating transcription. Thus, the selection of different poly(A) signals in a precursor-RNA can be used to regulate gene expression. For example, if the first poly(A) signal is ignored, a second poly(A) signal further downstream in a transcription unit may be used to incorporate novel exons into the mRNA. As an example of this type of regulation, the production of secreted or membrane-bound immunoglobulin M (IgM) is described. When a hematopoietic stem cell differentiates to a pre-B lymphocyte it produces IgM as an

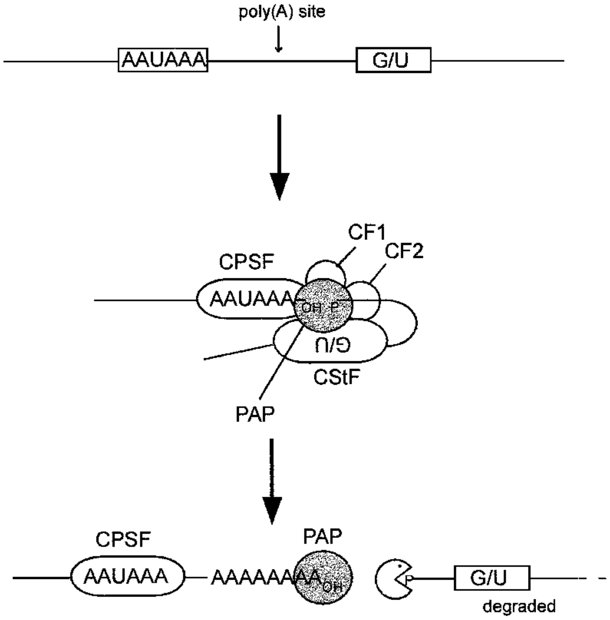

FIGURE 12 Model for cleavage and polyadenylation of a precursor-RNA in mammalian cells. The cleavage and polyadenylation specificity factor (CPSF) binds to the conserved AAUAAA sequence located 10–35 nucleotides upstream of the poly(A) site. The cleavage-stimulatory factor (CStF) binds to the GU-rich element located downstream of the poly(A) site. Binding of two cleavage factors (CF1 and CF II) and the poly(A) polymerase (PAP) then stimulates cleavage of the precursor-RNA. The PAP synthesizes the 200- to 250-nucleotide-long poly(A) tail, while the downstream RNA fragment is rapidly degraded.

antibody anchored to the plasma membrane. After binding to an antigen the lymphocyte undergoes differentiation and produces the same IgM molecule as a secreted protein. The shift from making a membrane-bound or secreted antibody is regulated at the level of alternative poly(A) site usage (Fig. 13). Thus, in the unstimulated B cell the first poly(A) signal is ignored and the downstream poly(A) signal is used as the default poly(A) site. The last two exons incorporated in the mRNA encode for a membrane-binding domain that anchors the antibody to the plasma membrane. After stimulation the upstream poly(A) signal is activated resulting in the processing of a mRNA that is translated to the same antibody but lacking the membrane-binding domain. Hence this antibody is secreted.

Also, a new translational reading frame, encoding for a completely different protein, may be produced by alternative poly(A) site usage. For example, production of calcitonin, which occurs in thyroid cells, and the calcitonin-related protein, which is produced in the brain, is regulated by tissue-specific poly(A) site usage.

D. Transcription and RNA Processing Are Coordinately Regulated Events

Recent experiments suggest that transcription and RNA processing are tightly coupled events. Thus, the RNA polymerase that assembles at the promoter has been shown to associate with factors required for polyadenylation and SR-related proteins that may partake in RNA splicing. The CTD tail of RNA polymerase II appears to function as a platform that recruits the RNA processing factors to the preinitiation complex. Therefore, the elongating RNA polymerase appears to have been loaded with the factors required for polyadenylation, and probably

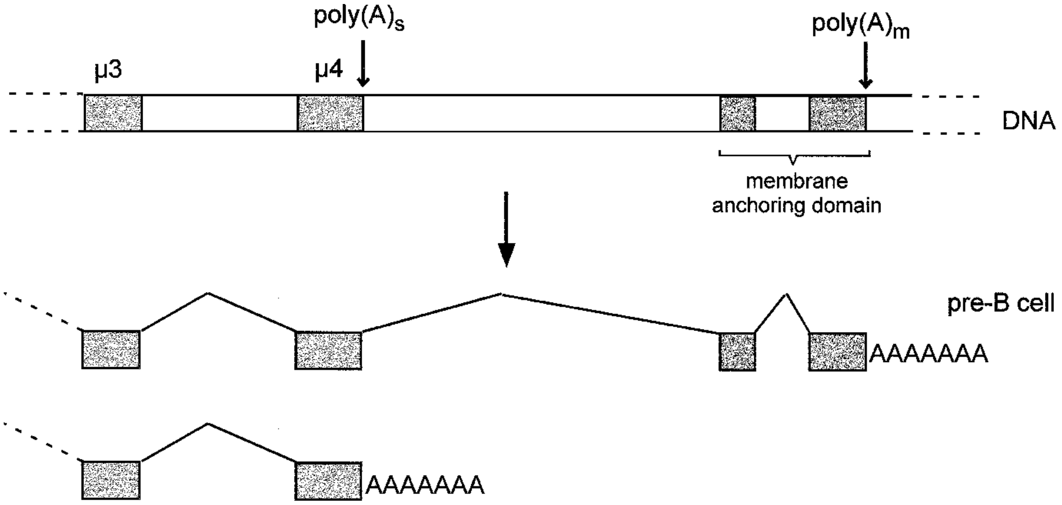

FIGURE 13 Alternative polyadenylation signals in the constant region of IgM yields heavy chains that are either membrane bound (pre-B cells) or secreted (plasma cells). Note that only the exon structure of the 3′ part of the IgM transcription unit is shown. The upstream region, which encodes for the antigen-binding domain, is identical in the two forms of IgM.

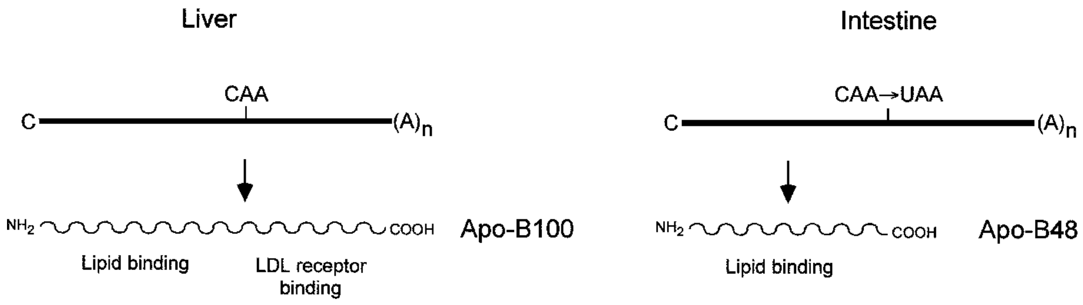

FIGURE 14 Editing of the Apo-B mRNA in intestinal cells. A CAA codon is edited to a UAA translational stop signal resulting in the production of a shorter protein (Apo-B48) corresponding to the amino-terminal half of the Apo-B100 protein expressed in liver cells.

deposits them at the polyadenylation signal used for 3′ end formation. Even more surprising, evidence has been presented suggesting that the composition of a promoter may specify alternative RNA splicing. Thus, the same gene under the transcriptional control of different promoters produces different types of alternatively spliced mRNAs. This finding suggests that enhancer binding transcription factors, in addition to stimulating recruitment of the RNA polymerase to the promoter, also may partake in the recruitment of selective RNA processing factors to the CTD tail. The future will tell whether alternative RNA splicing is regulated already at the level of transcription initiation.

E. Other Mechanisms of Posttranscriptional Regulation

Although this review has focused on control of gene expression at the level of synthesis and processing of mRNA, there are additional mechanisms that make significant contributions to gene expression. For example, a few genes in vertebrates have been shown to use RNA editing to produce different protein isoforms. The serum protein apolipoprotein B (Apo-B) is expressed in two forms. The Apo-B100 is expressed in hepatocytes, whereas a shorter polypeptide, Apo-B48, is expressed in intestinal epithelial cells. The cell-type-specific expression of Apo-B results from a posttranscriptional editing of the Apo-B mRNA in intestinal epithelial cells. Thus, a CAA codon encoding for the amino acid glutamate is converted to a UAA stop codon by cytosine deamination. As a result the Apo-B48 protein translated from the edited mRNA in the intestine differs from Apo-B100 by lacking the carboxy-terminus (Fig. 14). Both proteins bind to lipids. However, only the liver-specific Apo-B100 contains the carboxy-terminal domain required for binding to the low-density lipoprotein receptor, necessary for delivery of cholesterol to body tissues.

In nuclear genes RNA editing appears to be rare. Also, editing in such genes is restricted to modification of single nucleotides. In contrast, RNA editing in genes expressed in the mitochondria of protozoa, plants, and chloroplasts results in a more dramatic change of the mRNA sequence. Thus, a precursor-RNA may be edited such that more than 50% of the sequence in the mature mRNA is altered compared to the primary transcription product.

Gene expression is also regulated at other levels, such as nuclear to cytoplasmic transport of mRNA, translational efficiency of mRNA, RNA and protein stability, or protein modification. As is the case for transcriptional regulation, control of gene expression at the level of translation often occurs at the initiation step of the decoding process. Thus, not all mRNAs that reach the cytoplasm are used directly to synthesize protein. In fact, as much as 10% of genes in a eukaryotic cell may be regulated at the level of translation.

SEE ALSO THE FOLLOWING ARTICLES

Chromatin Structure and Modification • DNA Testing in Forensic Science • Enzyme Mechanisms • Hybridomas, Genetic Engineering of • Nucleic Acid Synthesis • Ribozymes • Translation of RNA to Protein

BIBLIOGRAPHY

Lodish, H., Berk, A., Zipursky, S. L., Matsudaria, P., Baltimore, D., and Darnell, J. (2000). "Molecular Cell Biology," Freeman and New York.

Latchman, D. S. (1998). "Eukaryotic Transcription Factors," Academic Press, San Diego, CA.

Gesteland, R. F., Cech, T. R., and Atkins, J. F. (1999). "The RNA World," 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Immunology—Autoimmunity

K. Michael Pollard
Eng M. Tan
Scripps Research Institute

I. Autoimmunity and Autoantibodies
II. Detection of Autoantibodies and Autoantigens
III. Perspectives

GLOSSARY

Antibody A protein product of a specific type of lymphoid cell (B cell) which combines with a specific molecular target, called an antigen.

Antigen A substance, usually protein, that combines with an antibody.

Autoantigen A constituent of the body that is recognized by its immune system. An antibody that recognizes an autoantigen is called an autoantibody.

Chromophore A chemical group capable of selective light absorption resulting in the coloration of certain organic compounds.

Neoantigen A new or different antigen.

IN NORMAL circumstances a healthy immune system is tolerant of the tissue constituents of the host but intolerant of nonhost (or "nonself") matter such as invading viruses, bacteria, or other pathogens. In autoimmunity the immune system mounts responses against tissue constituents of the host organism. This response is most commonly referred to as recognition of "self," as opposed to "nonself." The self/nonself paradigm is argued by many to be the basis for understanding how the immune system determines whether or not to respond to a suspected challenge. Considerable debate has occurred regarding the definition of self and non-self, with both "infection" and "danger" being suggested as possible discriminators. Irrespective of the final outcome of these discussions, it is clear that idiopathic autoimmunity constitutes a significant disruption in the mechanisms that regulate the immune systems' ability to discriminate at the molecular level. Autoimmunity can lead to devastating diseases such as systemic lupus erythematosus (SLE) and insulin-dependent diabetes mellitus (IDDM). These two clinical syndromes reflect the two major types of autoimmunity. SLE is representative of multisystem autoimmunity, in which disease processes may involve a number of organs of the body. IDDM is illustrative of organ-specific autoimmunity, which is more restrictive, targeting perhaps a single organ such as the pancreas. Advances in our understanding of the relationships between autoimmune diseases and the autoantibodies and autoantigens that characterize them have come about through the use of an ever-evolving armamentarium of techniques based in part on the physical sciences. This chapter is an attempt to describe the history, principles, and methodology of those techniques that have proven the most useful in the characterization of autoantibodies and autoantigens in both basic and applied research

arenas. To appreciate fully the significant roles these techniques have played it is necessary first to review briefly some of the fundamental features of autoimmunity and autoantibodies.

I. AUTOIMMUNITY AND AUTOANTIBODIES

Autoimmune responses constitute an attack by the immune system on the host. The responsible effector mechanisms appear to be no different from those used to combat exogenous agents and include soluble products such as antibodies (humoral immunity), as well as direct cell-to-cell interaction leading to specific cell-lysis (cell-mediated immunity). Autoantibodies are therefore defined as antibodies produced by the host which recognize cellular or tissue constituents of the host. No single mechanism has been described that can account for the diversity of autoimmune responses or the production of autoantibodies. Perhaps the most perplexing and challenging aspect of autoimmunity and autoantibody elicitation is the identification of the events involved in the initiation of the response. Although these early events are poorly understood for most autoimmune diseases, it is thought that an exogenous trigger may provide the first step in the initiation of some autoimmune responses. It is also uncertain how T and B cells, with receptors for autoantigen, emerge from primary lymphoid tissues, having escaped the regulatory mechanisms that normally delete them or keep them in check, and make their way to secondary lymphoid tissues where they can be activated to respond in an inappropriate manner. Studies involving transgenic mice expressing neoautoantigens suggest that possible mechanisms for emergence of autoractive cells include avoidance of apoptotic elimination, escape from tolerance induction, and reversal of an anergic state. Molecular identification of autoantigens, their presence in macromolecular complexes, the occurrence of autoantibodies to different components of the same complex, and the appearance of somatic mutations in the variable regions of autoantibodies have suggested that autoantigen drives the autoimmune response. These findings support the argument that activation of autoreactive cells occurs in secondary lymphoid tissues. It remains unclear how autoantigens, particularly intracellular autoantigens, are made available to the immune system and what molecular forms of these complex macromolecular structures interact with autoreactive lymphoid cells.

An important component in humoral autoimmune responses is the appearance of autoantibody-secreting B cells. The antibody secreted by a B cell is directed against a single region (or epitope) on an antigen. An autoantibody response can target a number of epitopes on any one antigen, clearly showing that multiple autoreactive B-cell clones are activated during an autoimmune response. In the systemic autoimmune diseases many autoantigens are complexes of nucleic acid and/or protein, and an autoimmune response may target several of the components of a complex. It is unknown whether the autoantibody responses to the components of a complex arise simultaneously, sequentially, independently, or through some interrelated mechanism. Immunization studies in mice have suggested that some autoantibody responses may arise sequentially, a process termed epitope spreading. Whether this mechanism is applicable to humans is under investigation.

In only a few instances have autoantibodies been shown to be the causative agents of pathogenesis (e.g., anti-acetylcholine receptor autoantibodies in myasthenis gravis, anti-thyroid stimulating hormone receptor autoantibodies in Graves' disease). It is noteworthy that these diseases are usually organ specific and that their autoantigens are extracellular or on the surface of cell membranes and therefore easily targeted by the immune system. In the non-organ-specific autoimmune disease systemic lupus erythematosus (SLE) anti-double-stranded DNA (dsDNA) autoantibodies have been shown to participate in pathogenic events by way of complexing with their cognate antigen to cause immune complex mediated inflammation. These examples show that in both organ-specific and systemic autoimmune diseases, *in vivo* deposition of autoantibody in tissues and organs has clinical significance, as it indicates sites of inflammation and possible pathological lesions. Detection of autoantibody/autoantigen deposits in organ-specific autoimmune diseases has particular significance because passive infusion of some organ-specific autoantibodies has been found directly to mediate pathological sequelea. In most autoimmune diseases, however, it has not been determined whether autoantibodies cause or contribute in any way to disease. It is possible that autoantibodies are an indicator of the primary event or a secondary consequence of the underlying clinical condition; autoantibodies have been described as potential "reporters" of disease mechanisms.

Diseases associated with autoantibodies can be divided into two broad groups: multisystem autoimmune diseases, in which autoantibodies react with common cellular components that appear to bear little relationship to the clinical syndrome, and organ-specific autoimmune diseases, in which autoantibodies have the ability to react with autoantigens from a particular organ or tissue. In both situations the specificity of the autoantibody can serve as a diagnostic marker (Table I). There are several features of the relationship between autoantibody specificity and diagnostic significance within the multisystem autoimmune diseases that bear consideration. Autoantigens in these

TABLE I Examples of the Clinical Diagnostic Specificity of Autoantibodies[a]

Autoantibody specificity[a]	Molecular specificity	Clinical association
Organ-specific autoimmune diseases		
Anti-acetylcholine receptor*	Acetylcholine receptor	Myasthenia gravis
Anti-TSh receptor*	TSH receptor	Graves' disease
Antithyroglobulin*	Thyroglobulin	Chronic thyroiditis
Anti-thyroid peroxidase*	Thyroid peroxidase	Chronic thyroiditis
Antimitochondria*	Pyruvate dehydrogenase complex	Primary biliary cirrhosis
Antikeratinoctye*	Desmoplakin I homologue	Bullous pemphigoid
Antikeratinoctye*	Desmoglien	Pemphigus foliaceus
Multisystem autoimmune diseases		
Anti-double-stranded DNA*	B form of DNA	SLE
Anti-Sm*	B, B′, D, and E proteins of U1, U2, and U4–U6 snRNP	SLE
Anti-nRNP	70-kDa, A, and C proteins of U1-snRNP	MCTD, SLE
Anti-SS-A/Ro	60- and 52-kDa proteins associated with the hY1–Y5 RNP complex	SS, neonatal lupus, SLE
Anti-SS-B/La	47-kDa phosphoprotein complexed with RNA polymerase III transcripts	SS, neonatal lupus, SLE
Anti-Jo-1*	Histidyl tRNA synthetase	Polymyositis
Antifibrillarin*	34-kDa protein of box C/D containing snoRNP (U3, U8, etc.)	Scleroderma
Anti-RNA polymerase 1*	Subunits of RNA polymerase 1 complex	Scleroderma
Anti-DNA topoisomerase 1 (anti-Scl-70)*	100-kDa DNA topoisomerase I	Scleroderma
Anticentromere*	Centromeric proteins CENP-A, -B, and -C	CREST (limited scleroderma)
cANCA	Serine proteinase (proteinase 3)	Wegener's vasculitis

[a] SLE, systemic lupus erythematosus; MCTD, mixed connective tissue disease; SS, Sjögren's syndrome; cANCA, cytoplasmic antineutrophil cytoplasmic antibody; TSH, thyroid-stimulating hormone; CREST, calcinosis, Raynaud's phenomenon, esophageal dysmotility, sclerodactyly, telangiectasia. Disease-specific diagnostic marker antibodies indicated by an asterisk.

diseases are components of macromolecular structures such as the nucleosome of chromatin and the small nuclear ribonucleoprotein (snRNP) particles of the spliceosome, among others. Autoantibodies to different components of the same macromolecular complex can be diagnostic for different clinical disorders. For example, the core proteins B, B′, D, and E, which are components of the U1, U2, and U4–U6 snRNPs and are antigenic targets in the anti-Smith antigen (Sm) response in SLE, are different from the U1 snRNP specific proteins of 70 kDa, A, and C, which are targets of the anti-nuclear RNP (nRNP) response in mixed connective tissue disease (MCTD; see Table I). It has also been shown that particular autoantibody responses are consistently associated with one another. The anti-Sm response, which is diagnostic of SLE, is commonly associated with the anti-nRNP response, but when the anti-nRNP response occurs in the absence of the anti-Sm response it can support the diagnosis of MCTD. These two observations suggest that the snRNP complexes responsible for the autoantibody response against the spliceosome in MCTD may differ from the snRNP complexes that produce the antispliceosome response in SLE. Other autoantibody responses demonstrate similar associations and restrictions. The anti-SS-A/Ro response (see Table I) frequently occurs alone in SLE but the anti-SS-B/La response in Sjögren's syndrome is almost always associated with the anti-SS-A/Ro response. Similarly the antichromatin response occurs alone in drug-induced lupus but is often associated with the anti-dsDNA response in idiopathic SLE. Autoantibody specificities may occur at different frequencies in a variety of diseases, and the resulting profile consisting of distinct groups of autoantibodies in different diseases can have diagnostic value. In some cases the grouping of autoantibody specificities, such as the preponderance of antinucleolar autoantibodies in scleroderma (Table I), provides intriguing but as yet little understood relationships with clinical diagnosis. Unlike SLE, in which a single patient may have multiple autoantibody specificities to a number of unrelated nuclear autoantigens (e.g., DNA, Sm, SS-A/Ro), scleroderma patients are less likely to have multiple autoantibody specificities to nucleolar autoantigens that are unrelated at the macromolecular level (i.e., not part of the same macromolecular complex).

TABLE II Examples of Subcellular Structures and Domains Recognized by Autoantibodies[a]

Autoantibody	Molecular specificity	Subcellular structure
Nuclear components		
Antichromatin	Nucleosomal and subnucleosomal complexes of histones and DNA	Chromatin
Anti-nuclear pore	210-kDa glycoprotein (gp210)	Nuclear pore
Antilamin	Nuclear lamins A, B, C	Nuclear lamina
Anticentromere	Centromere proteins (CENP) A, B, C, F	Centromere
Anti-p80 coilin	p80-coilin (80-kDa protein)	Coiled or cajal body
Anti-PIKA	p23- to 25-kDa proteins	Polymorphic interphase kayrosomal association (PIKA)
Anti-NuMA	238-kDa protein	Mitotic spindle apparatus
Nucleolar components		
Antifibrillarin	34-kDa fibrillarin	Dense fibrillar component of nucleolus
Anti-RNA polymerase 1	RNA polymerase 1	Fibrillari center of nucleolus
Anti-Pm-Scl	75- and 100-kDa proteins of the Pm–Scl complex	Granular component of nucleolus
Anti-NOR 90	90-kDa doublet of (human) upstream binding factor (hUBF)	Nucleolar organizer region (NOR)
Cytosolic components		
Antimitochondria	Pyruvate dehydrogenase complex	Mitochondria
Antiribosome	Ribosomal P proteins (P_0, P_1, P_2)	Ribosomes
Anti-Golgi	95- and 160-kDa golgins	Golgi apparatus
Antiendosome	180-kDa protein	Early endosomes
Antimicrosomal	Cytochrome P450 superfamily	Microsomes
cANCA	Serine proteinase (proteinase 3)	Lysosomes
Antimidbody	38-kDa protein	Midbody
Anti-centrosome/centriole	Pericentrin (48 kDa)	Centrosome/centriole

[a] NuMA, nuclear mitotic apparatus; Pm–Scl, polymyositis–scleroderma; cANCA, cytoplasmic antineutrophil cytoplasmic antibody.

The molecular spectrum of autoantigenic targets (see Tables I and II) together with their exquisite antigenic specificity has made autoantibodies valuable reagents in molecular and cellular biology. The most visually impressive demonstration of the usefulness of autoantibodies as biological probes is the indirect immunofluorescence (IIF) test. Using this technique (see Section II.A), an increasing number of autoantibody specificities are being identified that recognize cellular substructures and domains (Table II and Fig. 1). Autoantibodies against chromatin and DNA can be used to identify the cell nucleus. Other nuclear structures such as the nuclear lamina, which underlies the nuclear envelope, can be identified by antilamin autoantibodies as a ring-like fluorescence around the nucleus (Fig. 1a). The nucleolus and its subdomains can be identified by a variety of autoantibodies (Table II). Antifibrillarin autoantibodies, which recognize the highly conserved 34-kDa fibrillarin [a component of some small nucleolar RNP (snoRNP) particles], identify the dense fibrillar component. Autoantibodies to RNA polymerase I, although rare, have proven to be a useful marker for the fibrillar center of the nucleolus. A growing appreciation of the antigenic diversity of cellular constituents, together with improvements in fluorescent microscopy, has led to identification of autoantibodies reacting with a variety of subnuclear domains and compartments, some considerably smaller than the nucleolus. The coiled body, a small circular subnuclear structure originally described by the Spanish cytologist Santiago Ramon y Cajal in 1903, and now named the Cajal body, is an example. Cajal bodies can be identified using autoantibodies that react with p80 coilin (Fig. 1d), an 80-kDa protein highly enriched in Cajal bodies. Using other autoantibodies in colocalization studies, it has been found that Cajal bodies contain other proteins, including fibrillarin (previously thought to be restricted to the nucleolus and prenucleolar bodies). Autoantibodies have also been identified that react with subcellular structures other than the nucleus (Fig. 1). Prior knowledge of the existence and relative distributions of these subcellular organelles was instrumental in

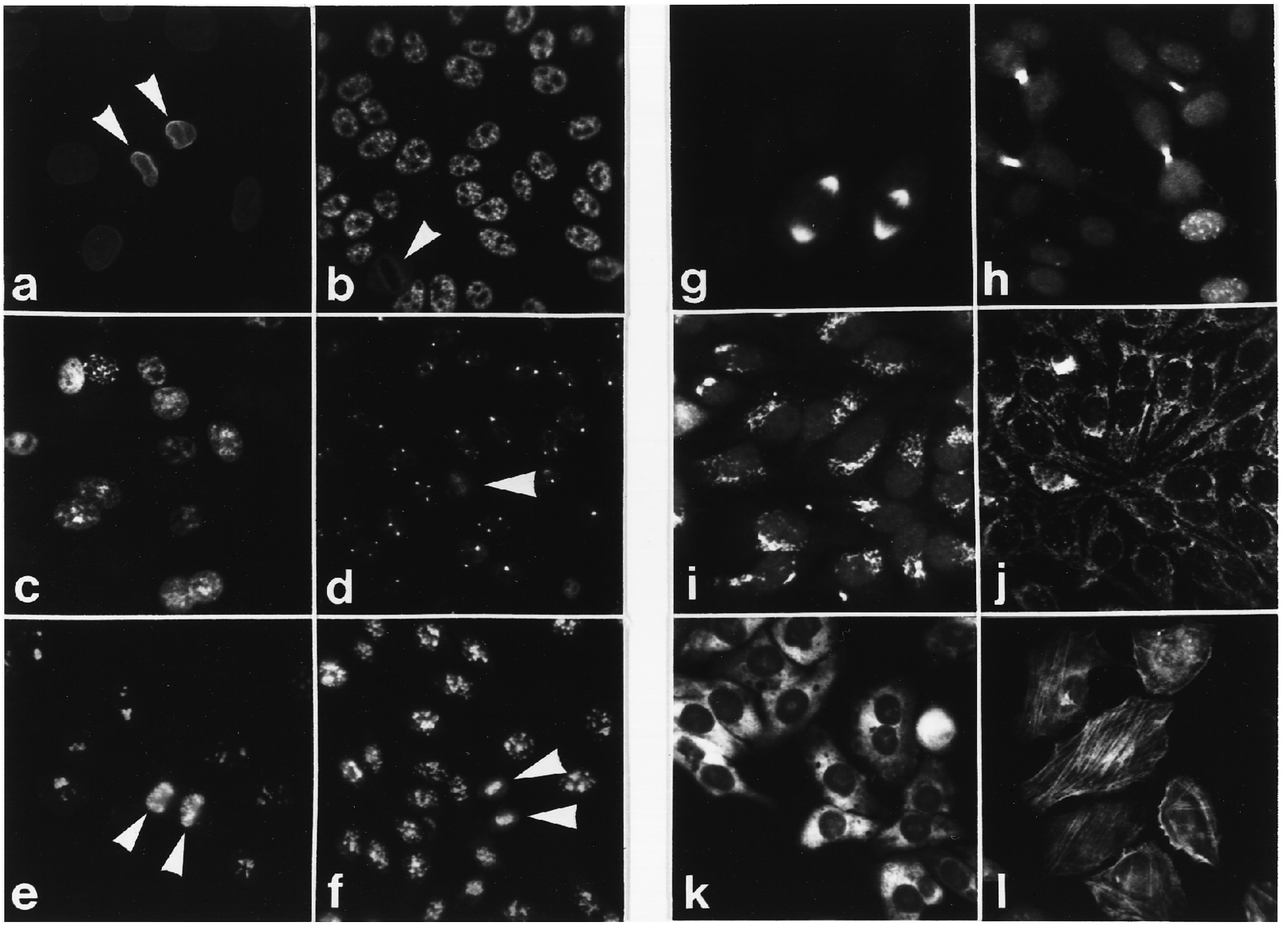

FIGURE 1 Immunofluorescence patterns produced by autoantibodies recognizing structural domains within the nucleus (a–f) and cytosol (g–l) of the cell. (a) Antinuclear lamin B_1 antibodies identify the periphery of the nucleus; arrowheads show the reformation of the nuclear envelope during late telophase. (b) Anti-Sm antibodies localize the U1, U2, and U4–U6 snRNP particles as a speckled pattern, but are absent from metaphase cells (arrowhead). (c) Anti-PCNA antibodies recognize the auxiliary protein of DNA polymerase delta during active DNA synthesis, producing different fluorescence patterns as cells progress through mitosis. (d) Anti-p80 coilin antibodies highlight subnuclear domains known as cajal or coiled bodies, which disappear during metaphase (arrowhead). (e) Antifibrillarin antibodies target the nucleolus, produce a characteristic clumpy pattern in interphase cells, and decorate the chromosomes from late metaphase until cell division (arrowheads). (f) Antibodies to centromeric proteins A, B, and C produce a discreet speckling of the interphase nucleus and identify the centromeric region of the dividing chromosomes during cell division (arrowheads). (g) Anti-mitotic spindle apparatus antibodies identify spindle poles and spindle fibers during cell division. (h) Antimidbody antibodies react with the bridge-like midbody that connects daughter cells following chromosome segregation but before cell separation. (i) Anti-Golgi complex antibodies decorate the Golgi apparatus, which in most cells is shown as a discreet accumulation of fluorescence in the cytoplasm. (j) Antimitochondrial antibodies demonstrate the presence of mitochondria throughout the cytoplasm; the discreet nuclear dots represent an additional autoantibody specificity in this serum unrelated to mitochondria. (k) Antiribosome antibodies produce a diffuse cytoplasmic staining pattern that spares the nucleus but may show weak nucleolar fluorescence, (l) Anticytoskeletal antibodies react with a variety of cytoskeletal components; in this case the antibody reacts with nonmuscle myosin. Original magnification: a–f and h–l, 350×; g, 700×.

identifying the structures recognized by these autoantibodies. Conversely autoantibodies, by virtue of their reactivity with individual autoantigens, have allowed cell and molecular biologists insight into the molecular constituents of these same subcellular organelles.

Autoantibodies can be used to study changes in the size, shape, and distribution of subcellular structures during the cell cycle, viral infection, mitogenesis, or any cellular response that results in alterations of subcellular constituents. For example, anti-lamin B1 autoantiodies can be

used to demonstrate the reformation of the nuclear lamina during telophase (Fig. 1a). Autoantibodies have also identified unexpected distributions of autoantigens, such as the distribution of the nucleolar protein fibrillarin to the outer surface of the chromosomes during cell division (Fig. 1e, arrowheads). The localization of some autoantigens during the cell cycle has aided in their identification. Detection of proliferating cell nuclear antigen (PCNA) in S-phase cells (Fig. 1c) suggested its involvement in DNA synthesis, while the distribution of speckles along the metaphase plate produced by other antibodies (Fig. 1f, arrowheads) was a significant contribution to their identification as autoantibodies against the centromeric proteins A, B, and C.

A feature of autoantibodies that underscores their uniqueness is their ability to recognize their target antigen not only from the host but also from a variety of species. The extent of this species cross-reactivity is dependent on the evolutionary conservation of the autoantigen and is related to the conservation of protein sequence. One example is the snoRNP protein fibrillarin. Using autoantibodies in a variety of techniques, this protein can be found in species as diverse as humans and the unicellular yeast *Saccharomyces cerevisiae*. cDNA cloning of fibrillarin has confirmed the expected high degree of conservation of the protein sequence.

Autoantibodies react with the conserved sequence and conformational elements of their cognate antigens; these features have made them useful regents in the cloning of cDNAs of expressed proteins from cDNA libraries from a variety of species. However, because of their reactivity with the human protein, they have been used primarily to clone the cDNAs and characterize the primary structures of numerous human cellular proteins. The diversity of the targets that have been exploited by this approach is illustrated in Tables I and II.

Elucidation of the structure of the autoantigens that are the targets of autoantibodies from systemic autoimmune diseases has revealed that many are functional macromolecular complexes involved in nucleic acid or protein synthesis. A distinguishing feature of many of these complexes of nucleic acid and/or protein is that autoantibodies do not recognize all the components of the complex. An extreme, but useful, example is the ribosome, which in eukaryotes may contain more than 70 proteins. However few of these proteins are recognized by autoantibodies, the major targets being the sequence-related P proteins (P_0, P_1, P_2). Nonetheless, the use of autoantibodies that identify specific components of such complexes has aided in identifying other subunits of these complexes, with profound consequences. Thus the initial identification of anti-Sm and anti-nRNP autoantibodies in SLE led to the observation that they recognize some of the protein components of the snRNP particles, fueling subsequent studies that showed the snRNPs as components of the spliceosome complex that functions in pre-mRNA splicing.

As the molecular and functional associations of autoantigens have become known, attempts to uncover the particular role of individual autoantigens have revealed that autoantibodies can directly inhibit the function of their cognate autoantigen. Although it remains to be determined, it seems likely that such inhibition reflects the involvement of conserved protein sequence or structure in functional activity. An increasing number of autoantibodies, many of unknown molecular specificity, recognize their autoantigen only in a particular functional state or phase of the cell cycle. Of the several examples known, the best characterized is PCNA, which is the auxillary protein of DNA polymerase delta and is recognized by autoantibodies only during mitosis, even though PCNA is present throughout the cell cycle. When a population of cells at different stages of the cell cycle is used in immunofluorescence anti-PCNA, autoantibodies produce varying degrees of fluorescence intensity, being negative for G_0 cells and highly positive for S-phase cells (Fig. 1c). These intriguing features of some autoantibodies have added new dimensions to their biological usefulness and have suggested that functionally active macromolecular complexes may play a role in the elucidation of autoantibody responses.

II. DETECTION OF AUTOANTIBODIES AND AUTOANTIGENS

The most commonly used methods for the detection and characterization of autoantibodies, whether in clinical medicine or molecular/cellular biology, fall under several broad biophysical areas and include fluorescent, enzymatic, and radiographic techniques. As described below some of these techniques have been specifically developed for antibody detection, while others have been borrowed and/or adapted from the biological/biophysical sciences. The methodology employed in these techniques has been described in limited detail to afford the reader the opportunity of understanding how the technique is put into practice at the laboratory bench, particularly in the study of autoimmunity.

A. Immunofluorescence

1. History

Characterization of the antigenic specificity of serum antibody by immunofluorescent methods dates to the early 1950s. In what was the forerunner of current technology, Weller and Coons, in 1954, used human serum to identify

viral antigens in tissue culture monolayers infected with varicella and herpes zoster. The human antibody bound to viral antigen in the infected cells was detected by a fluorochrome covalently linked to antibodies raised against human immunoglobulin. Although widely used for the detection of antibody against viral (or nonself) antigens, this application was soon followed by studies using the method to detect autoantibodies. Fluorescent anti-human globulin was used in 1957 by Friou and collegues, and Holborow and his co-workers, to demonstrate staining of the cell nucleus by serum from patients with systemic lupus erythematosus (SLE). These studies, using mouse and human tissue sections respectively, also revealed for the first time the species nonspecificity of the antinuclear "factor" (ANF), allowing tissue from a wide variety of sources to be used. Subsequent studies showed the ANF to be antibody, particularly immunoglobulin G (IgG), hence the current terminology of ANA (antinuclear antibody). Following closely behind these early studies came reports that the pattern of nuclear staining differed between patients with a single autoimmune disease as well as between patients with different autoimmune diseases. Thus homogeneous staining of the nucleus was more likely in SLE, while nucleolar staining was often found in the serum of patients with scleroderma. During this same period it was shown that patients with organ-specific autoimmune diseases had serum antibodies which reacted with antigens found in the organ targeted by the disease. Thus patients with Hashimoto's thyroiditis were shown to have serum antibody directed against antigens in the thyroid. The immunofluorescent test, due to the strong relationship between immunofluorescence "pattern" and autoantibody specificity, continues to be an important clinical test.

2. Principle

The principle of fluorescence is based upon the observation that certain substances adsorb radiation and become "excited" from a ground state of potential energy to the first excited state of the molecule. If the excited molecule is sufficiently stable, it will emit radiation rather than dissipate energy to the surrounding molecules as it returns to its ground state. The resulting emission is known as fluorescence and is almost always of a longer wavelength than the exciting radiation. This relationship between excitation and emission wavelengths is known as Stokes' law.

Epifluorescence is the most common method employed in fluorescence microscopy of antibodies. In this technique the excitatory radiation is passed through the objective lens *onto* the specimen, rather than *through* the specimen. This means that only reflected excitatory radiation needs to be filtered out to detect the emitted radiation. This is a significant advantage over transmission of light through the specimen, as significantly less irrelevant wavelength radiation needs to be filtered out. The filtering elements of an epifluorescence microscope consist of an excitation filter, an emission filter, and a dichroic mirror (Fig. 2). The excitation filter allows transmission of only those wavelengths that will excite the specific fluorochrome being used. The emission (or barrier) filter blocks transmission of the excitatory wavelength light but allows any fluorescence emitted to pass. The dichroic mirror is coated glass that is positioned 45° to the optical path of the microscope. The dichroic mirror functions as a beam splitter, reflecting the excitatory wavelength onto the specimen but allowing the emitted fluorescence to pass through to the eyepiece (Fig. 2). In most instances these three elements are housed together in a filter cube, with many microscopes able to hold two or more filter cubes. Each set of filters in a cube is a matched set and is restricted in its use to a small number of fluorochromes, usually one.

3. Method

Immunofluorescence is a relatively straightforward technique consisting of four major steps.

a. Preparation of cell or tissue substrates. A variety of cell and tissue substrates from numerous species have been used in immunofluorescent microscopy to characterize both autoantibodies and autoantigens. The most useful of these have been transformed mammalian cell lines grown in tissue culture. These cell lines contain the greatest variety of autoantigens and are particularly suited to the detection of autoantibodies in multisystem autoimmune diseases where tissue specificity of the autoantibody is not a confounding consideration. A primary concern in the preparation of cell or tissue substrates for immunofluorescence is that the antigenic integrity of the substrate be preserved during the fixation procedure necessary to stabilize the substrate for experimental use. Another important aspect of the fixation process is that it permeablizes the cell, thereby allowing antibody to enter and interact with its cognate antigen. Fixation usually involves organic compounds, such as ethanol and acetone. Determination of the optimal fixation conditions for the detection of specific antigens may require considerable experimentation.

b. Addition of primary antibody. Primary antibody may consist of an unknown specificity, such as that in a clinical sample being tested for the presence of autoantibodies, or a known specificity which is being used to determine the presence of the cognate autoantigen in a particular cell or tissue substrate.

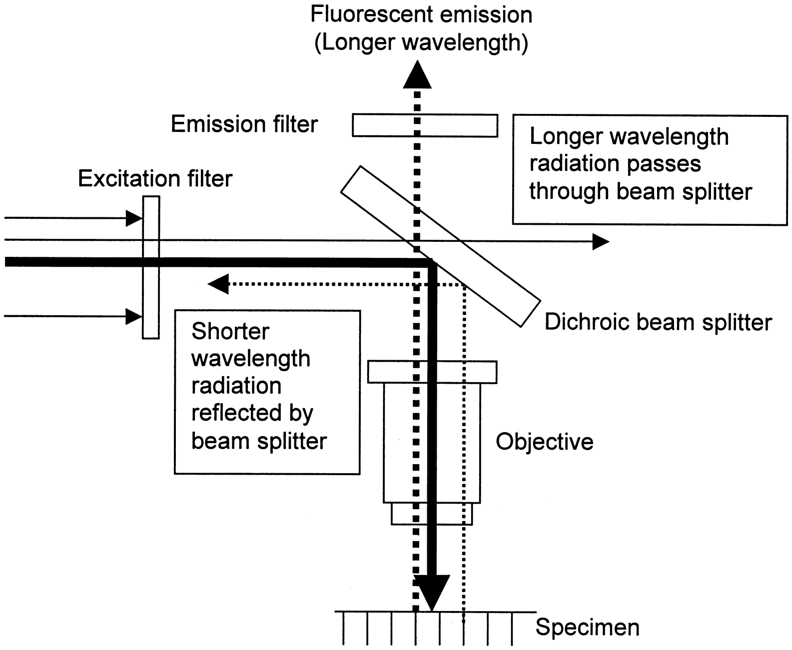

FIGURE 2 Function of filter cube elements in epifluorescence microscopy. The three elements of the filter cube, the excitation filter, emission filter, and dichroic beam splitting mirror, combine to allow selective use of narrow bandwidths of radiation in conjunction with specific fluorochromes so that fluorescence can be detected. The excitation filter allows only certain wavelengths to pass (thick black line), with the dichroic beam splitter reflecting an even narrower bandwidth toward the specimen. As described by Stokes' law this excitatory wavelength causes the fluorochrome conjugated to antibody to emit radiation of a higher wavelength (thick dashed black line), which passes through the dichroic beam splitter and is transmitted by the emission filter. This radiation is observed as a fluorescent image depicting the location in the specimen of the antigen bound by the fluorescently labeled antibody.

c. Addition of secondary antibody. The secondary (or detecting) antibody (antiserum) is immunoglobulin, usually highly purified, that is specific for the species from which the primary antibody originated. Thus if human serum is being tested for the presence of ANA, then the detecting antibody will be anti-human immunoglobulins. To be useful in immunofluorescence the secondary antibody is conjugated to a fluorochrome, most commonly fluorescein isothiocyanate (FITC) or rhodamine. However, a number of other fluorochrome are also used, including Alexa 488 and cyanine (Cy3) dyes, which produce more stable fluorescence.

d. Interpretation of results. Under optimal experimental conditions a serum containing an autoantibody will bind to its cognate antigen in the cell or tissue substrate, and this bound antibody will in turn be recognized by the fluorochrome-labeled secondary antibody. When viewed through a fluorescence microscope, the emitted fluorescence identifies the location of the antigen/autoantibody complex in the cell or tissue substrate. Success in interpreting the immunofluorescent "pattern" relies on many factors, including the optical properties of the microscope, the specificity and fluorescent label/protein ratio of the detecting antibody, and the experience of the observer. Considerable uncertainty in "pattern" recognition can be avoided through the use of control sera containing defined autoantibody specificities, such as those available through the Centers for Disease Control, Atlanta, Georgia.

B. ELISA

1. History

The first report of the use of ELISA (enzyme-linked immunosorbent assay) to measure antigen specific antibody was made by Engvall and Perlman in 1972, although the term ELISA had been used earlier by these workers to describe a competitive immunoassay for the quantitative measurement of antigen in solution. Since those early studies ELISA methodology has rapidly expanded to encompass a variety of techniques for the detection of antigen-specific and nonspecific antibodies, soluble and insoluble antigens, and cellular antigens, indeed any substance which can generate a specific antibody response. In autoimmunity the technique has become invaluable as a screening assay to detect autoantibodies to a variety of autoantigens, including proteins and nucleic acids. In

1990 Burlingame and Rubin described one of the most informative analyses of autoantibody reactivity as determined by ELISA. These investigators used subnucleosome structures as substrates in ELISA to demonstrate that complexes of DNA and histones (e.g., DNA complexed to a dimer of histone 2A and 2B) are better antigens than the individual components (e.g., histone 2B). These studies introduced an additional dimension to the concept that autoantibodies recognize conformational antigenic determinants by suggesting that the quarternary macromolecular structure may be the target of an autoantibody response. Molecular cloning of autoantigens and their expression and purification as recombinant proteins have significantly increased the usefulness of ELISA in identifying the specificity of autoantibodies in autoimmunity.

2. Principle

In a typical screening assay to detect an antigen-specific (auto)antibody, the antigen is adsorbed onto a solid support, usually a polystyrene microtiter plate. A source of soluble antibody (e.g., serum) is allowed to react with antigen, unbound material is washed away, and then an enzyme-conjugated antiimmunoglobulin is added. Unbound (excess) antibody conjugate is removed by washing, and the amount of antigen-specific antibody determined by the addition of an enzyme substrate and measurement of the enzyme-generated signal. The signal can be in the form of a color change as observed when alkaline phosphatase hydrolyzes *p*-nitrophenylphosphate to produce the yellow *p*-nitrophenolate. This color change can be measured using a spectrophotometer at 400 nm. Other means of generating signal include enzyme-generated fluorescence and luminescence.

3. Method

The majority of ELISA procedures that seek to measure the presence of autoantibodies employ direct adsorption of autoantigen onto a solid support such as a microtiter plate. The most important aspect of these assays is the purity of the antigen. If impure antigen is used, it may be difficult to define accurately the antigenic specificity of the autoantibody. For example, if the antigenic extract contains the nuclear snRNP particles, then both anti-Sm and anti nRNP autoantibodies will be detected. As these autoantibodies can define different clinical syndromes, any confusion in their detection is undesirable. On the other hand, the use of nucleosomes, rather than purified histone, is preferred for detection of antichromatin autoantibodies. Recent developments in ELISA include the use of purified recombinant autoantigens. Again, these antigenic preparations must be carefully characterized, particularly those from bacterial expression systems, as most individuals have antibodies to bacterial components and contamination of recombinant protein with bacterial antigens can lead to significant "false-positive" reactions.

C. Immunoblotting

1. History

The term immunoblotting essentially describes the transfer of antigenic material from one phase to another and the use of immunological reagents (i.e., antibody) to detect the transferred protein. The technique has also been called protein blotting and Western blotting, an attempt to distinguish the method from the related techniques of Southern and Northern blotting, which allow identification of DNA and RNA, respectively. The term immunoblotting is more applicable to the technique, particularly when used to characterize the antigenic specificity of autoantibodies. An early and elegant application of the procedure was described by Towbin and collegues in 1979. They electrophoretically transferred proteins, separated by polyacrylamide gel electrophoresis, onto a sheet of nitrocellulose (Fig. 3) and then detected the transferred protein using specific antibody that had been either radiolabeled or conjugated with a fluorochrome or an enzyme. This method exploited the ability of polyacrylamide gel electrophoresis to separate a mixture of proteins on the basis of their molecular weights and allowed that level of resolution to be transferred to nitrocellulose, where specific antibody could be used to identify individual proteins. This method was of immediate applicability to autoimmunity, as contemporary immunological techniques only allowed sera to be identified as having the same antigenic specificity, and did not allow identification of the specific antigen at the molecular level. Within a matter of a few years immunoblotting provided not only the molecular weight of many autoantigens, but also clues to the macromolecular complexity of multicomponent antigens such as the snRNP particles which contain the Sm and nRNP antigens. As with immunofluorescence, immunoblotting continues to be one of the major techniques used to characterize both autoantigens and autoantibodies.

2. Principle

Since the early descriptions of the method, immunoblotting has undergone considerable modification and many different techniques are in use today (see the Bibliography for additional reading). The description that follows is essentially the principle of electrophoretic transfer to nitrocellulose from polyacrylamide gels, the technique pioneered by Towbin, Staehelin, and Gordon in 1979. The

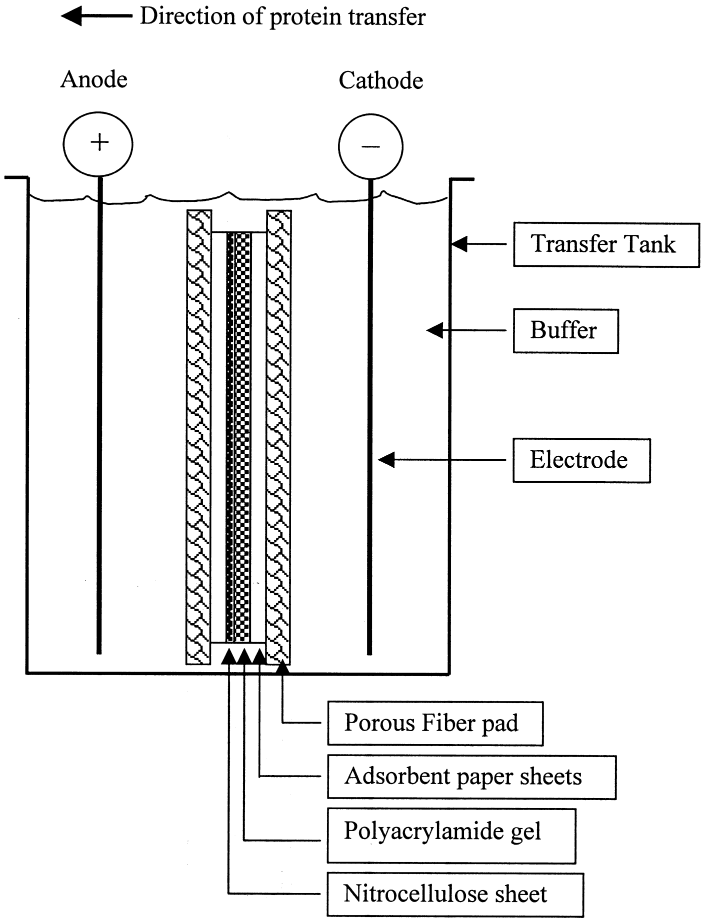

FIGURE 3 Diagrammatic representation of apparatus used in electrophoretic transfer of proteins from polyacrylamide gel to nitrocellulose. The transfer assembly consists of a "sandwich" made up of an outer porous fiber pad overlaid with several sheets of adsorbent paper, the polyacrylamide gel, a sheet of nitrocellulose, and then an additional layer of adsorbent paper sheets and another porous fiber pad. Assembly is done with all the components saturated with buffer to avoid air bubbles that may hinder protein transfer, particularly when between the acrylamide gel and the nitrocellulose sheet. The sandwich is held together by clamps or other firm support and immersed in a buffer-filled tank. Applying voltage results in the proteins being electrophoretically transferred from the polyacrylamide gel to the nitrocellulose. In the case of SDS-PAGE, the negatively charged proteins migrate toward the anode.

exact mechanism that allows proteins to bind to nitrocellulose is largely unknown, although hydrophobic forces may contribute. Under conditions of sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE), proteins are denatured with SDS and become negatively charged. The passing of an electric current through the polyacrylamide gel forces proteins to migrate toward the positive anode, with the proteins being resolved according to their molecular weight by virtue of the porosity of the gel. Transfer of the resolved proteins to nitrocellulose occurs according to essentially the same principle. Under an electric current the negatively charged proteins migrate out of the polycarylamide gel and onto the nitrocellulose (Fig. 3). An important feature of the technique is the absence of significant SDS in the transfer buffer, as SDS is known to reduce protein adsorption to nitrocellulose.

3. Method

An immunoblotting protocol consists of numerous steps, from SDS-PAGE resolution of a mixture of antigens to interpretation of the results (Fig. 4). For simplicity we consider here only those steps relating to SDS-PAGE resolution, electrophoretic transfer, and immunological detection of autoantigens. The source of protein for SDS-PAGE analysis should be one that contains the antigen or antigens of interest. One means of achieving this is to use a cellular substrate that has been found suitable for detection

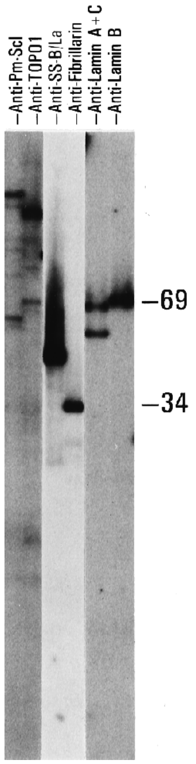

FIGURE 4 Immunoblot of rat liver nuclei using human autoimmune sera. Proteins in purified rat liver nuclei were resolved by SDS-PAGE, transferred to nitrocellulose, and then probed with human sera containing autoantibodies to a variety of different autoantigens. Anti-Pm-Scl serum blots a prominent band at 100 kDa; anti-DNA topoisomerase I (anti-TOPO1) blots a prominent band between 90 and 100 kDa; anti-SS-B/La reacts with a band at 48 kDa; antifibrillarin recognizes a band at 34 kDa; antibodies to nuclear lamins A and C recognize lamin A (higher molecular weight band) and the alternatively spliced lower molecular weight lamin C; antibodies to lamin B_1 react with a 69-kDa band. Numerical values to the right of the immunoblots represent protein molecular weight markers (kDa).

of the antigen using immunofluorescence. For characterization of human autoantibodies extracts from mammalian tissues (e.g., liver) or cell culture lines have been found to be suitable. If the antigen is only a minor component of the total cellular protein (e.g., the nucleolar protein fibrillarin), it may be necessary to purify the appropriate subcellular compartment (e.g., either the nucleus or the nucleolus) to enrich for the antigen of interest. Electrophoretic transfer of autoantigens to nitrocellulose, or other supports, is no more technically difficult than that of other proteins, however, immunological detection of transferred protein, using autoantibodies, can be challenging. Unlike antibodies raised by immunization, which are often very high in titer, autoantibody titers can vary considerably between sera. If immunoblotting is being used to detect the presence of an (auto)antigen, then a high-titer (auto)antiserum should be used. If the task is to screen a group of sera for autoantibodies to a particular autoantigen, then care should be taken to ensure that the level of antigen is sufficient and that appropriate "negative" controls are used so that "false-positive" reactions can be eliminated. Considerable uncertainty in recognition of the appropriate bands for particular autoantigens can be avoided through the use of control sera containing relatively monospecific autoantibody specificities, such as those available through the Centers for Disease Control, Atlanta, Georgia. Considerable attention should also be paid to the selection of the secondary antibody used to detect autoantibody bound to antigen immobilized on nitrocellulose. This reagent should be highly specific (e.g., affinity purified) and of a high titer. Visualization of the autoantigen/autoantibody complex by secondary antibody can be achieved in a variety of ways. The most sensitive methods include conjugation of secondary antibody with enzymes such as alkaline phosphatase and horseradish peroxidase. The addition of appropriate substrate allows enzyme-catalyzed colorimetric, fluorescent, or luminescent reactions that can be readily quantified.

D. Immunoprecipitation

1. History

Immunoprecipitation assays can take a variety of forms. Early assays made use of the "immunoprecipitin reaction," which, by mixing of increasing amounts of soluble antigen and antibody, allowed the formation of antigen–antibody complexes. A precipitate would form at the point of "equivalence," or when neither antigen nor antibody was in excess. This type of assay could also be performed in agar or agarose gels (e.g., Oucherlony precipitin test). However, these precipitin assays were cumbersome and required large amounts of antigen and antibody to establish optimal conditions for precipitation. The discovery that components of the bacterial wall (e.g., protein A of *Staphyloccus aureus*) could bind the Fc portion of IgG provided a convenient means of "precipitating" antigen–antibody complexes. Fixed and killed *S. aureus* could be used to adsorb IgG from serum and then mixed with antigen and the resulting antigen–antibody complexes isolated by centrifugation. Thus purified the properties of the antigen and/or antibody could be further examined without contamination from other component of complex mixtures of either antigen or antibody. Covalent coupling of protein A to Sepharose beads further improved the technique by removing contaminating bacterial antigens. Immunoprecipitation using protein A–Sepharose beads has been particularly useful in the characterization of the macromolecular structure of autoantigens. Using lysates from whole cells as antigen, and SDS-PAGE to resolve the precipitated

antigen, it has been found that autoantigens often comprise complexes of proteins and nucleic acids, such as the snRNPs of the spliceosome. Precipitation of such macromolecular complexes is also the major disadvantage of immunoprecipitation as it does not allow identification of individual antigenic components. This drawback can be overcome by using individual components in immunoprecipitation assays (Fig. 5) or by subjecting the immunoprecipitated complex to immmunoblotting techniques.

2. Principle

Immunoprecipitation is used in autoimmunity to help define autoantibody specificity, as well as to identify the components of the cognate autoantigen. As autoantibodies are predominantly of the IgG class, the most commonly used reagent for immunoprecipitation is protein A bound to Sepharose beads. Protein A interacts with the Fc portion of IgG in a reaction that is pH sensitive. The strongest interaction occurs in buffers that are neutral or slightly basic in pH, while acidic pH can be used to elute immunoglobulin. Not all subclasses of IgG bind to protein A; human IgG3 binds poorly, as does mouse IgG1. Protein A from *S. aureus* has five IgG binding sites, and protein A coupled to Sepharose beads binds at least two IgG molecules. Once an autoantibody-containing serum has been allowed to react with protein A–Sepharose, the unbound antibody is washed away and a source of antigen added to the autoantibody–protein A–Sepharose beads. Subsequent identification of the autoantigen is achieved by virtue of prior labeling of the protein and/or nucleic acid components. This is usually achieved by metabolic labeling of rapidly dividing cell clutures with a radioactive precursor such as [^{35}S]methionine for proteins or ^{32}Pi for nucleic acids. The immunoprecipitated antigen is subjected to polyacrylamide gel electrophoresis to resolve the components and autoradiography to visualize the radiolabeled components.

Immunprecipitation using extracts from whole cells may not allow identification of individual autoantigens, particularly if the autoantigen is a component of a macromolecular complex. In this case identification of the antigenic component can be achieved by using the radiolabeled product from the cDNA of the suspected antigen. An example of this antigen-specific immunoprecipitation assay is shown in Fig. 5.

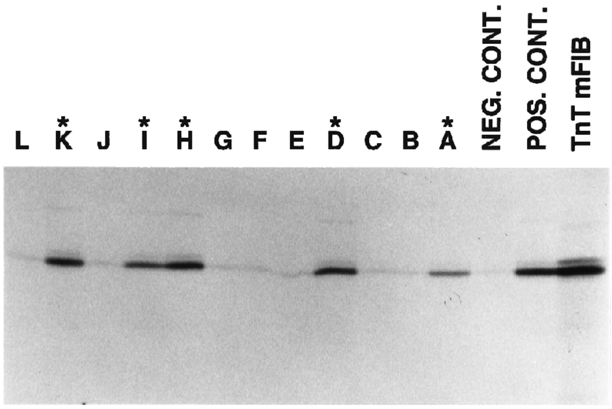

FIGURE 5 Immunoprecipitation of the autoantigen fibrillarin using autoantibodies. cDNA encoding mouse fibrillarin was radiolabeled with [^{35}S]methionine by *in vitro* transcription and translation (TnT mFIB). This protein was then used in a protein A–Sepharose bead immunoprecipitation assay to examine human sera (A–L) for antifibrillarin antibodies. Positive sera are identified by an asterisk. POS. CONT, immunoprecipitate from an antifibrillarin-positive serum; NEG. CONT., immunoprecipitate from an antifibrillarin-negative serum.

3. Method

As stated above immunoprecipitation of extracts from radiolabeled cells has allowed identification of many autoantigens as components of complexes of protein and nucleic acid (see Tables I and II). As with imunofluorescence and immunblotting, prior experimentation should be used to confirm the presence of the autoantigen of interest in the cell line serving as a source of autoantigens. Demonstration of the macromolecular structure of autoantigens requires considerable experimentation with different conditions of cell lysis and solubilization of cell extract. Conditions that are too stringent can lead to disruption of the complex, while mild conditions may not allow sufficient solubilization to release the complex from surrounding cellular constituents. As autoimmune sera can contain multiple autoantibody specificities, immunoprecipitation “patterns” revealed by autoradiography can be quite complex. Control sera containing defined autoantibody specificities, such as those available through the Centers for Disease Control, Atlanta, Georgia, should be used to help discern the “pattern” of molecular constituents of specific autoantigens.

III. PERSPECTIVES

Following the realization that autoantibody specificities can serve as diagnostic aids, considerable effort was, and continues to be, expended in developing appropriate test systems for use in research and clinical laboratories. The majority of these assays focus on the detection of autoantibody, although the target may be either a single antigen (e.g., immunoblot, ELISA, imunoprecipitation) or a complex mixture of antigens (e.g., immunofluorescence, ELISA, immunoprecipitation). Due to the diversity of autoantibody specificities, particularly in multisystem

autoimmune diseases such as SLE, immunofluorescence (using whole cells as the antigenic substrate) is the screening assay of choice. Apart from the use of cell culture lines rather than tissue sections, little change has been made to the principle of this assay since the 1950s. Fluorescence microscopy has seen several advances, including confocal microscopy and digitalization of images (including deconvolution), which have seen considerable use in basic research but have yet to see use in clinical screening of autoantibodies. During the last several decades all of the methods described have undergone numerous modifications. Most early modifications sought to improve safety by replacing the use of radioactive detecting reagents (e.g., ^{125}I-labeled secondary antibody) with other methods of detection such as enzyme-catalyzed technologies. These same modifications also improved the sensitivity and shortened the assay time by using colorimetric, fluorescent, or luminescent reactions, which are more readily detected and quantified than radioactive decay. As yet, high-throughput assays have not been applied to autoantibody detection, primarily because of the difficulty in screening an individual serum for the diverse spectrum of known autoantibody specificities. However, the availability of recombinant autoantigens and the increasing array of fluorescent or luminescent chromophores suggest that it should be possible to design assays capable of detecting and quantifying all autoantibodies in any serum using a complex mixture of autoantigens that are coupled to different chromophores.

SEE ALSO THE FOLLOWING ARTICLES

Cell Death (Apoptosis) • Chromatin Structure and Modification • Mammalian Cell Culture • Microanalytical Assays • Ribozymes

BIBLIOGRAPHY

Chan, E. K. L., and Pollard, K. M. (1997). Detection of autoantibodies to ribonucleoprotein particles by immunoblotting. *In* “Manual of Clinical Laboratory Immunology,” 5th ed. (N. R. Rose, E. Conway de Macario, J. D. Folds, H. C. Lane and R. M. Nakamura, eds.), pp. 928–934, American Society for Microbiology Press, Washington, DC.

Coligan, J. E., Kruisbeek, A. M., Margulies, D. H., Shevach, E. M., and Strober, W. (eds.) (2000). “Current Protocols in Immunology,” John Wiley and Sons, New York.

Krapf, A. R., von Mühlen, C. A., Krapf, F. E., Nakamura, R. M., and Tan, E. M. (1996). “Atlas of Immunofluorescent Autoantibodies,” Urban and Schwarzenberg, Munich.

Peter, J. B., and Shoenfeld, Y. (eds.). (1996). “Autoantibodies,” Elsevier, Amsterdam.

Spector, D. L., Goldman, R. D., and Leinwand, L. A. (eds.) (1997). “Cells: A Laboratory Manual,” Cold Springs Harbor Laboratory Press, Cold Springs Harbor, NY.

Ribozymes

Alessandra Poggi
John J. Rossi
Beckman Research Institute of the City of Hope

I. Introduction
II. Group I and Group II Introns
III. Ribonuclease P
IV. Self-Cleaving RNAs
V. Ribozymes as Tools and Gene Therapy Agents
VI. Conclusions

GLOSSARY

Cis The same strand of RNA.
Functional genomics Identifying the function of genetic sequences.
Gene therapy Supplying therapeutic genes or genetic elements to treat disease.
Genotype The genetic makeup of an organism.
Introns Sequences interrupting the coding sequence of a gene or RNA transcript.
Phenotype The result of gene expression.
Ribozymes RNA molecules with enzymatic activities.
Trans A different strand of RNA.

I. INTRODUCTION

Ribozymes are RNA molecules capable of acting as enzymes even in the complete absence of proteins. They have the catalytic activity of breaking and/or forming covalent bonds with extraordinary specificity, accelerating the rate of those reactions. They were simultaneously discovered in a group I self-splicing intron of the Tetrahymena pre-rRNA and in the ribonuclease (RNase) P enzyme purified from *Escherichia coli.* Today there are many additions to the ribozyme world, including members of the group I and II introns, the self-cleaving domains of the genomes of viroids, virusoids, hepatitis D, and a Neurospora satellite RNA. It has been proposed that the 23S RNA component of the ribosome may function as a ribozyme in translation, and the U6 snRNA in the spliceosome. Ribozymes occur naturally, but can also be artificially engineered and synthesized to target specific sequences in cis or trans. New biochemical activities are being developed using *in vitro* selection protocols as well. Ribozymes can easily be manipulated to act on novel substrates. These custom-designed RNAs have great potential as therapeutic agents and are becoming a powerful tool for molecular biologists.

The discovery of catalytic RNA molecules has revolutionized views on the origins of life. RNA molecules, once thought to be primarily passive carriers of genetic information, can carry out some functions previously only thought to be catalyzed by proteins, indicating that RNA can confer not only a genotype (as in many RNA viruses),

but also a phenotype. The RNA catalyzed reactions include self-cleavage, or trans-cleavage reactions, ligation, and trans-splicing. These observations have led to speculation that RNA might have been an early self-replicating molecule in the prebiotic world. Evidence supporting this notion comes from the fact that group I introns can exhibit RNA polymerase-like activities under certain conditions. In addition, the catalytic core of group I introns shares homology with several small satellite RNAs associated with plant viruses, which are also homologous to the human hepatitis delta virus (HDV), suggesting a common and ancient origin.

II. GROUP I AND GROUP II INTRONS

Introns are noncoding sequences that interrupt parts of genes. When a pre-mRNA is transcribed from the gene, introns need to be removed to give rise to mature messengers that will become a template for protein synthesis. Introns are removed by a process of cleavage-ligation called splicing. Generally, splicing requires a multicomplex of proteins and RNA. When introns were first discovered in nuclear genes, it was noted that all of their DNA sequences began with a GT and ended with an AG dinucleotide (Chambon's rule). Certain introns isolated from ribosomal or organellar genes, however, did not follow this simple rule, and DNA sequence comparisons led to the classification of these introns as either group I or group II on the basis of phylogenetically conserved sequence homologies and secondary structures. Some examples of group I and II introns are capable of self-splicing *in vitro* in the absence of protein. Those catalytic RNAs produce 5′-phosphate and 3′-OH termini on the reaction products.

A. Group I Introns

Group I introns are widely distributed in fungal mitochondria, chloroplasts, rRNA genes of protists, T-even phages, and the genomes of eubacteria. Group I intron self-splicing (*in vitro*) in the absence of proteins was first observed for the intervening sequence (IVS, intron) of the nuclear 26S rRNA gene in *Tetrahymena thermophila*.

Group I splicing proceeds by two consecutive transesterification reactions. These reactions are initiated by a nucleophilic attack by the 3′-hydroxyl of a guanosine (or a phosphorylated derivative: GMP, GDP, or GTP) at the phosphodiester bond between the 5′-exon and the intron (5′-splice site). The new 3′-hydroxyl group of the 5′-exon then initiates a second nucleophilic attack, this time on the phosphodiester bond between the 3′-exon and the intron (the 3′-splice site). This results in ligation of the exons and excision of the intron.

Despite all the evidence for self-splicing *in vitro*, it is clear that splicing *in vivo* requires protein factors. Even the Tetrahymena IVS, which at low levels of Mg^{2+} splices efficiently *in vitro*, is splicing at a rate of about 50-fold less than the level estimated for splicing *in vivo*. Proteins therefore aid in the folding of these complex RNAs to allow the self-splicing reaction to occur.

Self-splicing is, by definition, an intramolecular event, and the intron is therefore not acting as a true enzyme. However, the catalytic activity found within the conserved core, with a small deletion, can be dissociated into distinct active enzyme and substrate molecules. Cleavage at the 5′- and 3′-splice sites of group I introns can also occur slowly in the absence of a guanosine cofactor, due to the sensitivity of these sites to base hydrolysis which generates cleaved products consistent with the splicing reaction (3′-OH and 5′-P) but unusual for the hydrolysis reaction. The rate of this type of hydrolysis at the splice sites is much greater than expected (10-fold higher), implying that the folded RNA structure influences the susceptibility of certain phosphodiester bonds to alkaline hydrolysis.

Shortened versions of the Tetrahymena IVS (L-19 IVS and L-21 ScaI IVS) have been shown to be true enzymes *in vitro*, for example, as a restriction endoribonuclease and as a template-dependent polymerase.

B. Group II Introns

Group II introns present a relatively restricted distribution; they have been found in plant and fungal mtDNAs and comprise the majority of the introns in chloroplasts. It has also been demonstrated that some members of the group II introns can self-splice *in vitro*. The unimolecular reaction was shown to be Mg^{2+} dependent, requiring spermidine, having a temperature optimum of 45°C, and having a pH optimum of between 6.5 and 8.5. They differ from group I introns by the structure of their catalytic core and the intermediate and end products of splicing which involve a lariat structure. Again, the reaction consists of two transesterifications. Group II introns splice by way of two successive phosphate transfer reactions. In the first step, the 2′-OH group of an intramolecular branch point adenosine attacks the phosphodiester bond at the 5′-splice site (creating a 2′,5′-bond), producing the free 5′-exon and a splicing intermediate, the intron-3′-exon. The second step involves cleavage at the 3′-splice site by the 3′-OH of the 5′-exon. Simultaneously, the exons are ligated and the intron lariat, with a 2′,5′-phosphodiester bond, is released. Base-pairing interactions between sequences known as the exon binding site (EBS) and the intron binding site (IBS) hold the splice sites in close proximity. This ability of group II introns to specifically bind the 5′-exon has been exploited to encourage the intron to

catalyze reactions on exogenous substrates. The ability of the group II introns to bind the 5′-exon specifically has been exploited to encourage the IVS to catalyze reactions on exogenous substrates. These introns can be engineered to insert into target RNAs in trans in a reversal of the splicing reaction, thereby making them useful for site-specific gene inactivation or site-specific integration of therapeutic genes.

III. RIBONUCLEASE P

Ribonuclease P (RNase P) is a ubiquitous endoribonuclease that processes the 5′-end of precursor tRNA molecules, producing a 5′-phosphate and 3′-OH termini on the cleavage products. RNase P consists of both protein and RNA components, and it was shown that the catalyst was the RNA moiety. As with the catalytic introns, a divalent cation is required as cofactor. RNase P is unique among naturally occurring ribozymes in that it binds and cleaves free substrate molecules; all other characterized ribozymes act in cis. This natural trans-activity makes RNase P an obvious candidate for development as a therapeutic agent. Another feature that distinguishes RNase P from all other ribozymes is that it does not involve Watson–Crick base-pairing between the catalytic RNA and substrate for substrate recognition. Much effort therefore has been directed toward elucidating the biochemistry and substrate specificity of the RNase P cleavage reaction. One of the first aspects to be analyzed was the role of the protein subunit in the cleavage reaction. Comparisons of the kinetic aspects of the *B. subtilis* RNA-dependent cleavage reaction performed under various ionic conditions have demonstrated that high ionic strength and addition of the protein subunit have similar effects on the kinetics of cleavage. This may indicate that the protein subunit acts to disperse the charge repulsions between the RNase P and precursor tRNA substrate RNAs. Because many RNase P RNAs are not functional in the absence of protein *in vitro*, another possible role for the protein is to help the RNA moiety to fold into the proper conformation. As with most RNA-processing enzymes, the exact substrate requirements in terms of sequence and secondary structure are not well understood. It has been shown that mature tRNA can compete for binding, suggesting that most of the binding energy comes from mature tRNA. Analysis of pre-tRNA deletion mutants showed that only the amino acceptor stem and the T loop and stem, which form a single coaxially stacked helix, are required for cleavage by RNase P, suggesting that any hairpin structure can be cleaved by RNase P provided that a single-stranded NCCA trinucleotide is present at the 3′-side of the hairpin. Forster and Altman asked whether a double-stranded RNA substrate composed of two separate RNA molecules was also a suitable cleavage substrate. Mixing two short complementary oligoribonucleotides containing a 3′-proximal NCCA sequence resulted in cleavage of the target RNA at the predicted site. It seems, then, that any RNA can be cleaved by endogenous RNase P if an external guide sequence (EGS) containing a single-stranded NCCA at its 3′-end is provided to hybridize with the chosen target.

IV. SELF-CLEAVING RNAS

One category of intramolecular RNA catalysis is that which produces a 2′,3′-cyclic phosphate and 5′-OH terminus on the reaction products. A number of small plant pathogenic RNAs (viroids, satellite RNAs, and virusoids), a transcript from a Neurospora mitochondrial DNA plasmid, and the animal HDV undergo a self-cleavage reaction *in vitro* in the absence of protein. The reactions require neutral pH and Mg^{2+}. It is thought that the self-cleavage reaction is an integral part of their *in vivo* rolling circle mechanism of replication. These self-cleaving RNAs can be subdivided into groups depending on the sequence and secondary structure formed around the cleavage site.

A. Hammerhead Ribozymes

This group of RNAs shares a two-dimensional structural motif known as the hammerhead, which has been shown to be sufficient to direct site-specific cleavage. The hammerhead structure consists of three base-paired stems which flank the susceptible phosphodiester bond, and two single-stranded regions, which are highly conserved in sequence. Extensive mutagenesis has revealed the important nucleotides and functional groups for efficient catalysis. The hammerhead cleavage domain has been split into two or three independent RNAs, and trans-cleavage has been demonstrated *in vitro*. Haseloff and Gerlach proposed a model whereby the hammerhead domain is separated such that the substrate RNA contains just the cleavage site, and the ribozyme contains the other conserved nucleotides of the catalytic core. Mutagenesis has revealed that the target site can be any NUH sequence where H = A, C, and N is any nucleotide. The sequence of the arms of the ribozyme aligns the catalytic core to the target site via complementary base-pairing, Analysis has also allowed determination of the minimum core sequence required for catalytic activity. The ability to cleave the RNA and thereby inhibit the expression of a specific gene selectively has two main applications: as a surrogate genetic tool for molecular biology and the inactivation of gene transcripts *in vivo*, as antiviral agents, for example.

B. Hairpin Ribozyme

A second small catalytic domain is the hairpin structure, which has four helical domains and five loops. Two helices of the hairpin domain form between the substrate and ribozyme, and this allows the design and specificity of binding for trans-acting hairpin ribozymes. The hairpin ribozyme has a more complicated substrate requirement than the hammerhead ribozyme, but generally can cleave to the 5′-side of any GUC.

C. Hepatitis Delta Virus

HDV genomic and antigenomic RNAs contain a self-cleavage site hypothesized to function during rolling circle replication of this satellite virus. Like the plant pathogens, the sites in HDV are postulated to have a related secondary structure, three models of which have been proposed: cloverleaf, pseudoknot, and axehead, none of which is similar to the catalytic domains previously described. Like the other ribozyme motifs, the HDV ribozymes require a divalent cation, and cleavage results in products with 2′,3′-cyclic phosphate and 5′-OH termini. Investigations of trans-cleavage with the HDV ribozyme have not advanced to those of the hammerhead or hairpin ribozymes.

D. Neurospora Mitochondrial VS RNA

The Neurospora mitochondrial VS RNA, a single-stranded circular RNA of 881 nt, shares some features of the self-catalytic RNAs of HDV, group I introns, and some plant viral satellite RNAs. Although VS RNA can be depicted as having a secondary structure like group I introns, it is missing essential base-pairing regions, the cleavage site is in a different position, and the termini produced are 2′,3′-cyclic phosphate and 5′-OH. Like the hammerhead ribozymes, the VS RNA requires divalent cations for cleavage *in vitro*. The catalytic core of Neurospora VS RNA has been shown to consist of 154 nt.

V. RIBOZYMES AS TOOLS AND GENE THERAPY AGENTS

A. Ribozyme Delivery

Whatever type of ribozyme is chosen, it must be introduced into its target cell. Two general mechanisms exist for introducing catalytic RNA molecules into cells: exogenous delivery of the preformed ribozyme and endogenous expression from a transcriptional unit. Preformed ribozymes can be delivered into cells using liposomes, electroporation, or microinjection. Efforts have been made to overcome the lack of stability of introduced RNAs by using modified nucleotides, 2′-fluoro- and 2′-amino-, or 2′-O-allyl- and 2′-O-methyl-, mixed DNA/RNA molecules, or by the addition of terminal sequences (such as the bacteriophage T7 transcriptional terminator) at the Y-end of the RNA to protect against cellular nucleases. Endogenous expression has been achieved by inserting ribozyme sequences into the untranslated regions of genes transcribed by RNA polymerase II (pol II), which have strong promoters, such as the CMV promoter. Ribozymes have also been inserted into the anticodon loop of tRNA, transcribed by RNA polymerase III (pol III), or have been inserted into nonessential loop structures of small nuclear RNAs and a small nucleolar RNA. They have been shown to be functional in both plant and animal cells and in transgenic plants and animals.

Successful use of ribozymes to knockdown target gene expression is dependent on a number of factors, including target site selection as well as ribozyme gene delivery, expression, stability, and intracellular localization. The two types of ribozymes that have been used most extensively for knockdown studies are the hairpin and hammerhead ribozymes. Not all target sites for these ribozymes are accessible for cleavage: secondary structures, binding of proteins and nucleic acids, and other factors influence intracellular ribozyme efficacy. Computer-assisted RNA folding predictions and *in vitro* cleavage analyses are not necessarily predictive of intracellular *or vivo* activity, and the best ribozyme target sites often must be determined empirically *in vivo*. Alternative strategies utilizing cell extracts with native mRNAs have also proven useful for determining accessible ribozyme binding sites.

Ribozymes can be introduced into cells as genes by transfection or viral vector transduction, or as chemically synthesized molecules stabilized with various base substitutions and 3′ and 5′ modifications. Intracellular delivery, in this case, is typically achieved by some form of carrier molecule, such as cationic lipids. Depending on the application, either delivery method has its advantages and disadvantages. Delivery of the ribozyme genes can provide stable intracellular expression and the promoter choice can allow cell- or tissue-specific expression. Delivery of synthetic ribozymes allows short-term, high-level availability, making it easier to use as a "drug;" however, ribozyme stability and pharmacokinetics may present significant challenges.

Stable intracellular expression of transcriptionally active ribozymes can be achieved by viral vector-mediated delivery. Currently, retroviral vectors are the most commonly used in cell culture, primary cells, and in transgenic animals. Retroviral vectors have the advantage of stable integration into a dividing host cell genome, and the

absence of any viral gene expression reduces the chance of an immune response in animals. In addition, retroviruses can be easily pseudo-typed with a variety of envelope proteins to broaden or restrict host cell tropism, thus adding an additional level of cellular targeting for ribozyme gene delivery. Adenoviral vectors can be produced at high titers and provide very efficient transduction, but they do not integrate into the host genome and, consequently, expression of the transgenes is only transient in actively dividing cells. Other viral delivery systems are actively being pursued, such as the adeno-associated virus, alphaviruses, and lentiviruses. Adeno-associated virus is attractive as a small, nonpathogenic virus that can stably integrate into the host genome. An alphavirus system, using recombinant Semliki Forest virus, provides high transduction efficiencies of mammalian cells along with cytoplasmic ribozyme expression.

B. Ribozyme Gene Expression

A number of viral promoters (e.g., cytomegalovirus, retroviruses) utilizing RNA pol II have yielded high intracellular expression levels of ribozymes. Furthermore, selectable markers such as antibiotic resistance or cell surface proteins can be coexpressed with the ribozyme to allow monitoring of ribozyme expression. Embedding ribozymes in a long mRNA transcript can potentially have a negative impact on ribozyme localization and even folding. One possible solution for this problem is to engineer self-cleaving ribozymes on either side of the functional ribozyme so that it can be released from the nascent transcript.

RNA pol III promoters, including tRNA, adenovirus VA1, U1, and U6 small nuclear RNA derivatives, have allowed ubiquitous and very high intracellular levels of ribozyme expression. Several creative permutations, such as the combination of two promoters driving the same ribozyme gene or the placement of the ribozyme expression cassette within the U3 region of a Long terminal repeat, to produce a "double-copy" retroviral vector led to further increases in ribozyme expression. Finally, multiple ribozymes can be expressed within the same RNA transcript or from multiple promoter-ribozyme cassettes.

Intracellular localization of the ribozyme transcript is another important parameter. Specific strategies to colocalize ribozymes with their target RNA have been developed to maximize intracellular ribozyme activity. These strategies take advantage of the RNA sequences flanking the ribozyme. Certain RNA sequences direct specific subcellular localization, making it possible to tightly colocalize ribozyme transcripts with their target RNAs.

As the use of ribozymes progresses from cell culture systems to animal models, additional control over ribozyme expression will be required. Ribozyme expression can be restricted to specific organs or cell types through the use of tissue-specific promoters. This has been done successfully in tissue culture using the tyrosinase promoter, which is exclusively expressed in melanocytes. In another example, transgenic mice were created that carried a ribozyme gene driven by the insulin promoter, only expressed in the pancreatic beta-cell islets. Alternatively, inducible promoters, such as those regulated by tetracycline, or steroid hormones, have shown utility both in cell culture and in animals, allowing ribozyme expression to be turned on and off at will.

C. Ribozyme Applications

Ribozymes have been applied as antiviral agents, treatments for cancer and genetic disorders, and as tools for pathway elucidation and target validation. Ribozymes are unique in that they can inactivate specific gene expression, and thus can be used to identify the function of a protein or the role of a gene in a functional cascade. This application, called target validation, is critical for both basic biological research and drug development. Compared to other means of target validation (e.g., transgenic or knockout animals), ribozymes offer specificity and ease of design and usage.

Initial uses of ribozymes focused on antivirals, primarily for the treatment of HIV. Viruses that go through a genomic RNA intermediate in their replication cycle, such as HIV, hepatitis B virus, and hepatitis C virus, are attractive targets because a single species of ribozyme can target both viral genomic RNA and mRNAs. Ribozymes have also been widely used to target cellular genes, including those aberrantly expressed in cancers. One early ribozyme target was the bcr–abl fusion transcript created from the Philadelphia chromosome associated with chronic myelogenous leukemia. This chromosome is characterized by a translocation that results in the expression of a transforming bcr–abl fusion protein. In this case, ribozymes have been designed to specifically target the fusion mRNA and not the normal bcr or abl mRNAs. Ribozymes have also been designed to specifically target the mutant ras gene while sparing the normal homolog. Ribozymes targeting overexpressed HER-2/neu in breast carcinoma cells effectively reduced the tumorigenicity of these cells in mice.

In addition to directly targeting oncogenes, ribozymes have also been applied more indirectly as anticancer therapies. For example, ribozymes targeting the multiple drug resistance-1 or fos mRNAs in cancer cell lines effectively made the cells more sensitive to chemotherapeutic agents. Alternatively, a ribozyme targeting bcl-2 triggered

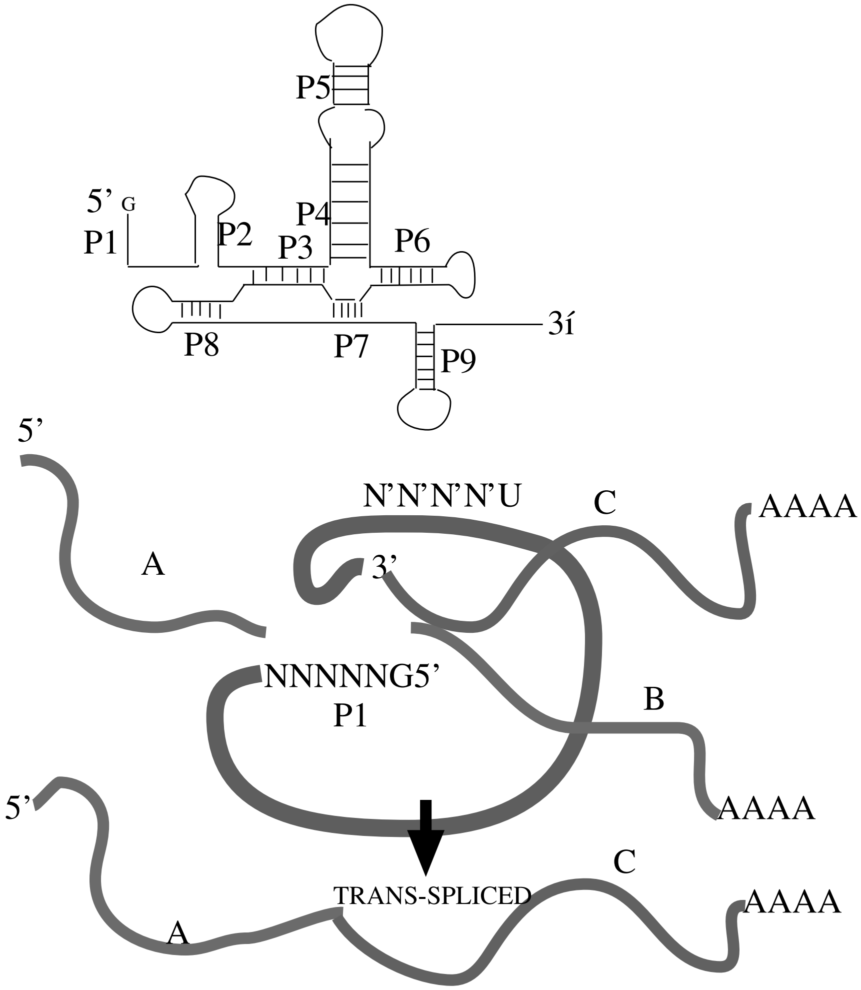

FIGURE 1 The group I intron. The top of this figure depicts the generalized secondary structure of the group I intron. This intron catalyzes the self-splicing of transcripts and can trans-splice as well, as depicted in the bottom portion of the figure. The P1 region forms a helix with the target RNA via Watson–Crick base-pairing, allowing the attached exon C to be trans-spliced with A.

apoptosis in these cells. Factors required for metastasis are also attractive targets for ribozymes. Ribozymes targeted against CAPL/mts, matrix metalloproteinase-9, pleiotrophin, and VLA-6 integrin all reduced the metastatic potential of the respective tumor cells in mice. Angiogenesis is also an important target for cancer therapy, and has been blocked in mice by ribozymes targeting fibroblast growth factor binding protein and pleiotrophin. Ribozyme-based therapies have also been tested in animals to inhibit other proliferative disorders, such as coronary artery restenosis.

Heritable and spontaneous genetic disorders represent additional applications for therapeutic ribozymes targeting cellular genes. These include the beta-amyloid peptide precursor mRNA involved in Alzheimer’s disease, and an autosomal-dominant point mutation in the rhodopsin mRNA that gives rise to photoreceptor degeneration and retinitis pigmentosa.

D. Functional Genomics and Target Validation

Ribozymes can be used to inactivate specific gene expression, and thereby can be used to help identify the function of a protein or the role of a gene in a functional biochemical pathway. Target validation is an increasingly important tool in basic biological research as well as in drug development. With the recent completion of the human genome sequencing initiative, there are tens of thousands of transcriptomes that have no assigned function. Ribozymes provide a facile and highly specific tool for interfering with the expression of these transcripts to monitor their biological function.

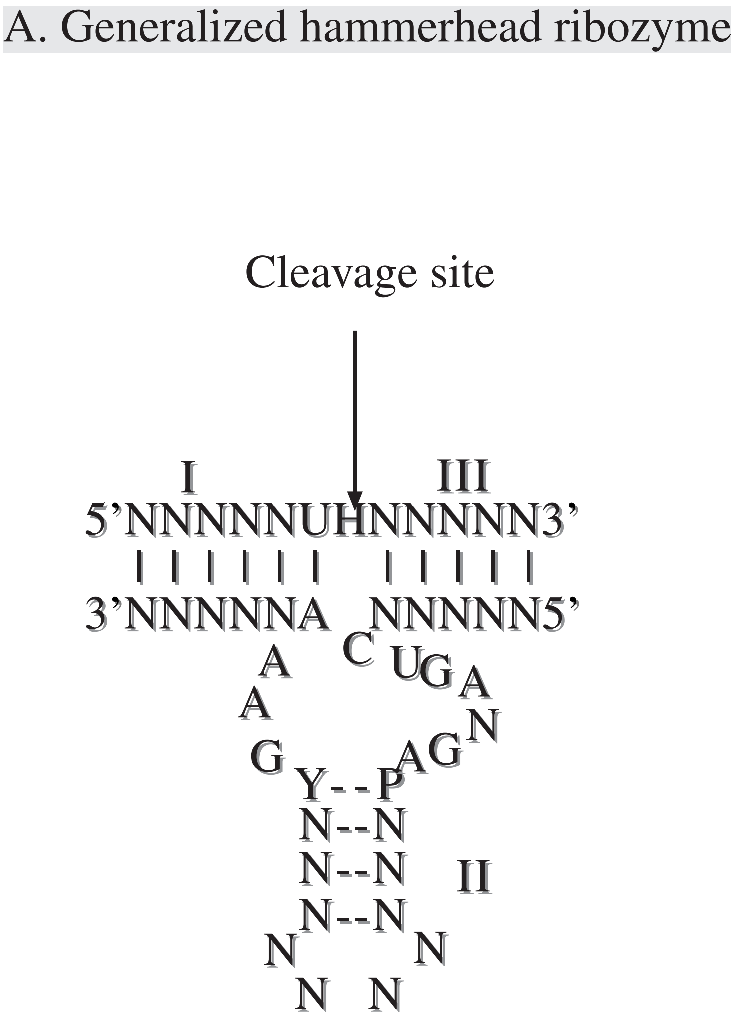

FIGURE 2 Generalized hammerhead ribozyme. The ribozyme catalytic core is flanked by stems I, II and III. Cleavage is directed after the H (A, C or U). The N’s represent any nucleotide, Y = pyrimidine, and P = purine. Other bases are A = adenosine, C = cytosine, G = guanosine, and U = uracil.

Ribozyme mediated target validation can also be used to identify specific member(s) of a protein family involved in a specific phenotype. Carefully designed ribozymes can selectively knockdown expression of each protein in a gene family.

E. Transgenic Animals Expressing Ribozymes

Therapeutics and target validation studies will certainly be tested in animals. Ribozymes have been used in transgenic mice to create disease models such as diabetes by selectively downregulating the hexokinase mRNA in pancreatic islets. In this case, the ribozyme expression was under the control of the insulin promoter, and was therefore only expressed in the pancreatic beta cells. Retroviral delivery of ribozymes targeted against neuregulin-1 in a chick blastoderm resulted in the same embryonic lethal phenotype as a gene knockout. Localized retroviral delivery of the same ribozyme later in development allowed dissection of the neuregulin biochemical pathway.

F. Ribozyme Mediated RNA Repair

A novel therapeutic application of ribozymes exploits the trans-splicing activity of the Tetrahymena ribozyme. This ribozyme has been used to repair defective mRNAs by trans-splicing onto these RNAs a functional sequence. These ribozymes are designed to bind and cleave the target RNAs 5′ of the undesired mutation. Since the ribozyme in this case is an intron, it is engineered to carry with it the correct RNA sequence as the 3′-exon. Following cleavage of the mutant target RNA, the ribozyme catalyzes ligation of wild-type sequence onto the cleaved transcript. This was successfully demonstrated with the correction of a

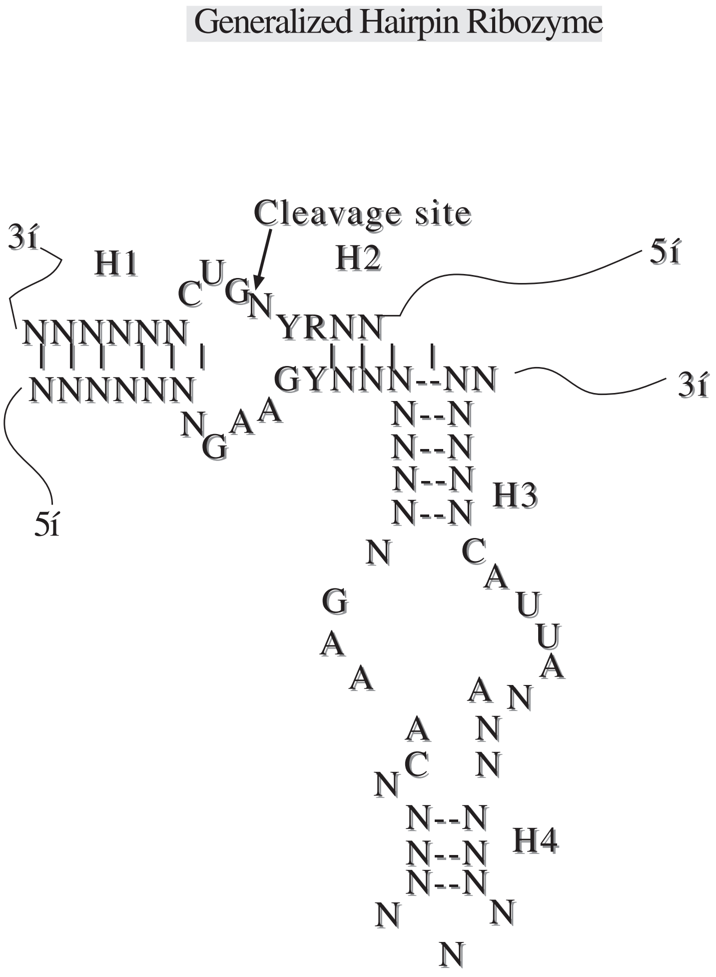

FIGURE 3 Generalized hairpin ribozyme. The target RNA (upper sequence) base pairs with the ribozyme (lower sequence) by Watson–Crick base-pairing. Cleavage occurs at the site indicated. The helices formed by the ribozyme and substrate are designated as H1–H4, as indicated. N = any nucleotide, Y = pyrimidine, P = purine. Other bases are A = adenosine, C = cytosine, G = guanosine, and U = uracil.

mutant lacZ transcript, first in bacteria and subsequently in correction of a sickle cell message in erythroid cells.

G. Ribozyme Based Gene Discovery

The necessity of ribozymes to selectively base-pair with the target RNAs prior to catalysis has been exploited to discover new gene functions involved in a particular phenotype. This protocol has been termed "inverse genomics." A combinatorial ribozyme library has been introduced into cultured cells, and selection or scoring for a particular phenotype has led to the identification of several new gene functions. A hairpin ribozyme library containing roughly 2×10^7 different ribozymes was stably transduced into tumor cells in culture and cells able to grow in soft agar were identified. Ribozymes isolated from these cells were sequenced, and the complement to the ribozyme pairing arms was determined. The identity of the target revealed a novel tumor suppressor. This approach has potential applications in many systems (therapeutic or otherwise) where a phenotype is of interest but the genes involved are unknown.

H. Ribozyme Evolution

The discovery of the ribozyme sparked new debate on the "RNA world" hypothesis, where all biological processes were carried out by RNA-based enzymes. Since then, RNA evolution has been forced *in vitro* to come up with RNA enzymes capable of carrying out a wide variety of biochemical reactions, as far-reaching as carbon–carbon bond and peptide bond formation. *In vitro* RNA evolution

has been used to create RNA-cleaving Rzs with smaller catalytic domains, DNA-cleaving ribozymes, and new catalytic motifs. Even RNA-cleaving DNAzymes have been generated through *in vitro* evolution. These "evolved" enzymes exemplify the power of *in vitro* evolution and will no doubt find many applications.

VI. CONCLUSIONS

The transformation of ribozyme sequences from naturally occurring, cis-cleaving molecules to target-specific, trans-cleaving reagents has stimulated a great deal of interest in their potential applications. Ribozymes targeting viral genes are now in clinical evaluation; ribozymes targeting cellular genes are moving into transgenic animals; and the use of ribozymes is expanding into RNA evolution, mRNA repair, and gene discovery.

For ribozymes to become generally useful surrogate genetic tools and realistic therapeutic agents, several obstacles need first to be overcome. These obstacles are the efficient delivery to a high percentage of the cell population, efficient expression of the ribozyme from a vector or intracellular ribozyme concentration, colocalization of the ribozyme with the target, specificity of ribozyme for the desired mRNA, and an enhancement of ribozyme-mediated substrate turnover. Despite these reservations, results with ribozymes so far look promising, particularly in the HIV-1 studies. As our knowledge of RNA structure, secondary and tertiary, increases, we will be able to target RNAs more rationally, which may help with the problems of specificity. At the same time, the understanding of the physical localization of RNA in cells and its tracking as it moves from the nucleus to cytoplasm will also help in ensuring colocalization of the ribozyme and target. Modifications of the ribozymes, for example, the 2′-ribose with various agents such as methyl, allyl, fluoro, and amino groups, increases the stability to nucleases quite dramatically. Similarly, chimeric DNA–RNA ribozymes increase the stability. The efficiency of delivery to cells with viral vectors or liposomes is also continually improving. These molecules must retain their catalytic potential, reach an accessible site on the substrate, and effectively impact on the steady-state levels of target molecules to be useful as either surrogate genetic tools or therapeutic agents. Great progress has been made in all of these areas and should allow extensive use of the highly specific reagents for down-regulating expression of target RNAs.

SEE ALSO THE FOLLOWING ARTICLES

Biomaterials, Synthesis, Fabrication, and Applications • Cell Death (Apoptosis) • Gene Expression, Regulation of • Immunology-Autoimmunity • Mammalian Cell Culture • Translation of RNA to Protein

BIBLIOGRAPHY

Castanotto, D., Rossi, J. J., and Deshler, J. O. (1992). "Biological and Functional Aspects of Catalytic RNAs," *Crit. Rev. Eukaryot. Gene Expr.* **2,** 331.

Cech, T. R. (1990). "Self-Splicing of Group I Introns," *Annu. Rev. Biochem.* **59,** 543.

Doherty, E. A., and Doudna, J. A. (2000). "Ribozyme Structures and Mechanisms," *Annu. Rev. Biochem.* **69,** 597.

Fedor, M. J. (2000). "Structure and Function of the Hairpin Ribozyme," *J. Mol. Biol.* **297,** 269.

Hauswirth, W. W., and Lewin, A. S. (2000). "Ribozyme Uses in Retinal Gene Therapy," *Progr. Retin. Eye Res.* **19,** 689.

James, H. A. (2000). "Therapeutic Potential of Ribozymes in Haematological Disorders," *Expert Opin. Investig. Drugs* **9,** 1009.

Kurz, J. C., and Fierke, C. A. (2000). "Ribonuclease P: A Ribonucleoprotein Enzyme," *Curr. Opin. Chem. Biol.* **4,** 553.

Rossi, J. J. (1999). "Ribozymes, Genomics and Therapeutics," *Chem. Biol.* **6,** R33.

Rossi, J. J. (2000). "Ribozyme Therapy for HIV Infection," *Adv. Drug. Deliv. Rev.* **44,** 71.

Szostak, J. W. (1997). "In Vitro Selection and Directed Evolution," *Harvey Lect.* **93,** 95.

Watanabe, T., and Sullenger, B. A. (2000). "RNA Repair: A Novel Approach to Gene Therapy," *Adv. Drug Deliv. Rev.* **44,** 109.

Translation of RNA to Protein

R. A. Cox
H. R. V. Arnstein
National Institute for Medical Research

I. Introduction
II. mRNA Structure and the Genetic Code
III. Transfer RNA
IV. Ribosome Structure and Function in Translation
V. Translational Control of Gene Expression
VI. Concluding Remarks

GLOSSARY

Anticodon Three consecutive bases in tRNA that bind to a specific mRNA codon by complementary antiparallel base-pairing.

Antiparallel base-pairing Pairing through specific hydrogen bonds between base residues of two polynucleotide chains or two segments of a single chain with phosphodiesterbonds running in the $5' \rightarrow 3'$ direction in one chain or segment and in the $3' \rightarrow 5'$ direction in the other. In DNA and RNA the hydrogen bonds are usually formed between complementary base pairs (A with either T or U, and G with C).

Elongation The stepwise addition of amino acids to the carboxyl terminus of a growing polypeptide chain.

Initiation A multistep reaction between ribosomal subunits, charged initiator transfer RNA, and messenger RNA that results in apposition of the ribosome-bound initiator Met-tRNA with an AUG initiator codon in mRNA. In this position, the ribosome is poised to form the first peptide bond.

Polarity The asymmetry of a polymer such as a polynucleotide or polypeptide. In DNA, the two strands have opposite polarity; that is, they run in opposite directions ($5' \rightarrow 3'$ and $3' \rightarrow 5'$). The polarity of a polypeptide is defined as running from the N-terminus to the C-terminus.

Reading frame One of three possible ways of translating groups of three nucleotides in mRNA. The appropriate reading frame is determined by the initiation codon.

Template strand The strand of the DNA double helix that is used as a template for transcription of RNA. It has a base sequence complementary to the RNA transcript.

Termination The end of polypeptide synthesis, which is signaled by a codon for which there is no corresponding aminoacyl-tRNA. When the ribosome reaches a termination codon in the mRNA, the polypeptide is released and the ribosome-mRNA-tRNA complex dissociates.

Translation The stepwise synthesis of a polypeptide with an amino acid sequence determined by the nucleotide sequence of the mRNA coding region. The genetic code relates each amino acid to a group of three

consecutive nucleotides termed a *codon*. Decoding of mRNA takes place in the $5' \rightarrow 3'$ direction, and the polypeptide is synthesized from the amino to the carboxyl terminus.

Translocation The stepwise advance of a ribosome along mRNA, one codon at a time, with simultaneous transfer of peptidyl-tRNA from the A site to the P site of the ribosome.

PROTEINS are essential to the structure and function of living cells. The assembly of polypeptide chains from amino acids and their subsequent modifications, leading to the final three-dimensional protein structure, are exceptionally complex processes; many components are involved and much of the cell's energy is utilized. Each peptide bond requires the expenditure of four high-energy phosphate bonds. This value excludes the energy used for initiation and release of the polypeptide chains and the cost of synthesizing and processing mRNA. The linear amino acid sequence of a protein is encoded within the gene as a linear deoxyribonucleotide sequence. Early steps in the biosynthesis of a protein include transcription of the gene and appropriate processing of the transcript leading to the production of mature messenger RNA (mRNA). We describe the mechanisms involved in translating mRNA to produce a polypeptide chain which has the amino acid sequence specified by the gene.

I. INTRODUCTION

A gene or cistron is defined as the region of DNA that is transcribed into a functional RNA. The transcript functions either as such (e.g., tRNA, rRNA, snRNA) or as a messenger (mRNA), which, after processing or editing as required, normally codes for one or more polypeptide chains in the translation process. A polynucleotide such as RNA is an asymmetrical polymer assembled from nucleoside triphosphates by a stepwise mechanism linking the $5'$ position of one nucleotide by a phosphate bridge to the $3'$ position of the adjacent nucleotide. In the finished polynucleotide chain the first nucleotide residue has a $5'$ position which is not linked to another nucleotide, whereas the last nucleotide has an unlinked $3'$ position. Thus, polynucleotide synthesis proceeds from the $5'$ to the $3'$ terminus and the polymer is said to have a $5'$ to $3'$ polarity. Usually, linear RNA sequences are written with the $5'$ terminus on the left and the $3'$ terminus on the right (Fig. 1A). Within the RNA chain some bases may form antiparallel base pairs (Fig. 1B).

The polypeptide chains of proteins are also asymmetrical polymers in which the amino acid residues are linked by peptide bonds between their alpha amino and carboxyl groups (Fig. 2), leaving a free alpha amino group at one end (the amino terminus) of the polymer and a free alpha carboxyl group at the opposite end (the carboxyl terminus). The significance of the polarity of RNA and proteins will become evident when the process of protein biosynthesis is explained below.

The genetic information stored in DNA is not usable directly for making proteins. Rather, it must be copied into a primary RNA transcript containing the mRNA by an enzymatic transcription of segments of DNA containing the genes. In prokaryotes, the primary transcript is also the messenger; that is, it can be used directly in polypeptide synthesis. In contrast, in eukaryotes the primary transcript is often much larger in size than the mature mRNA and requires extensive processing involving the excision of intervening and other noncoding sequences.

Messenger RNA serves as the template for protein synthesis; that is, the linear nucleotide sequence of the mRNA dictates the amino acid sequence of the polypeptide encoded originally by the gene. Conventionally, gene and mRNA nucleotide sequences are written in the $5'$ to $3'$ direction, which corresponds to the direction in which mRNA is decoded during polypeptide synthesis: The mRNA is read in the $5'$ to $3'$ direction, and the polypeptide is synthesized from the amino- toward the carboxyl-terminus.

The mechanism whereby RNA is translated into protein is complex, and the cell devotes considerable resources to the translational machinery. The components include 20 different amino acids, transfer RNAs, aminoacyl-tRNA synthetases, ribosomes, and a number of protein factors which cycle on and off the ribosomes and facilitate various steps in initiation of translation, elongation of the nascent polypeptide chain, and termination of synthesis with release of the completed polypeptide from the ribosome. The process depends on a supply of energy provided by ATP and GTP. The rate of protein synthesis is typically in the range of 6 (immature red blood cells of the rabbit) to 20 (*Escherichia coli* growing optimally) peptide bonds per sec. at 37°C.

II. mRNA STRUCTURE AND THE GENETIC CODE

A. Structure

The sequence information of a gene is copied (transcribed) into the nucleotide sequence of RNA from the complementary strand of DNA, called the template strand. The primary transcript is a single strand of RNA, which

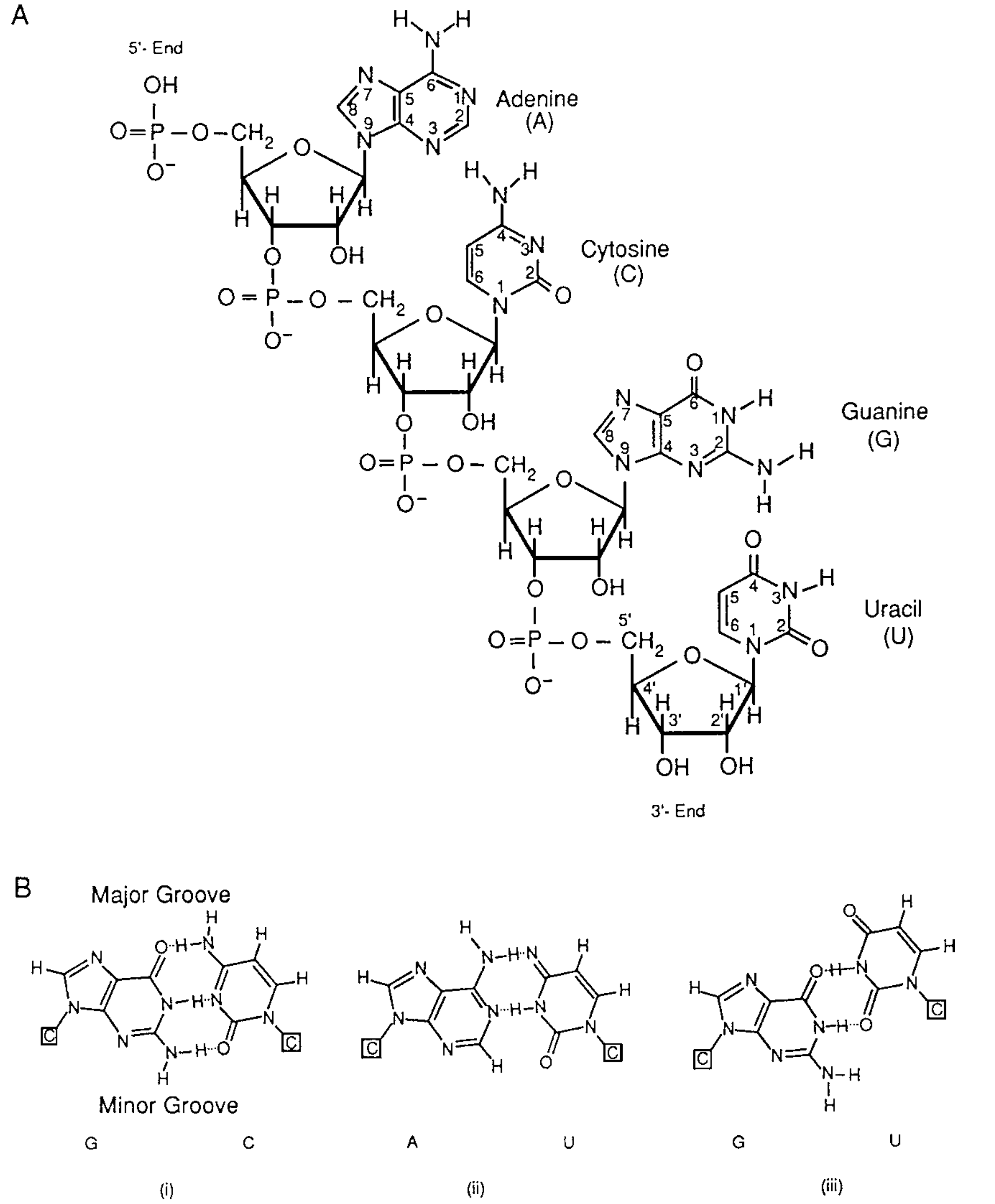

FIGURE 1 Structural elements of ribonucleic acids. (A) Primary structure indicating the numbering system for purines, pyrimidines, and ribose. (B) Base-pairing interactions commonly found in RNA: (i) G + C base pair; (ii) A + U base pair; (iii) G + U base pair. Pairs i and ii are Watson and Crick base pairs; iii is a special type of base pairing found in intramolecular bihelical regions. The minor groove is on the side of the base pair with the glycosidic bond of which the carbon atom C1′ of ribose is boxed. Note that DNA contains base pairs of types i and ii only but with thymine (5-methyluracil) in place of uracil. [From Arnstein, H. R. V., and Cox, R. A. (1992). "Protein Biosynthesis," Oxford University Press, Oxford. With permission.]

is a faithful copy of the *other* strand of DNA with substitution of U residues in place of T residues found in DNA. Sometimes, the primary transcript is altered, as described below, before it functions as mRNA; in these cases, the original unmodified transcript is the precursor or pre-mRNA. Usually, mRNAs have nontranslated sequences at the 5′ and 3′ ends in addition to the coding domain. These noncoding sequences sometimes affect the efficiency of translation and the stability of mRNA. The structures of typical mRNAs are shown in Fig. 3. The decoding process

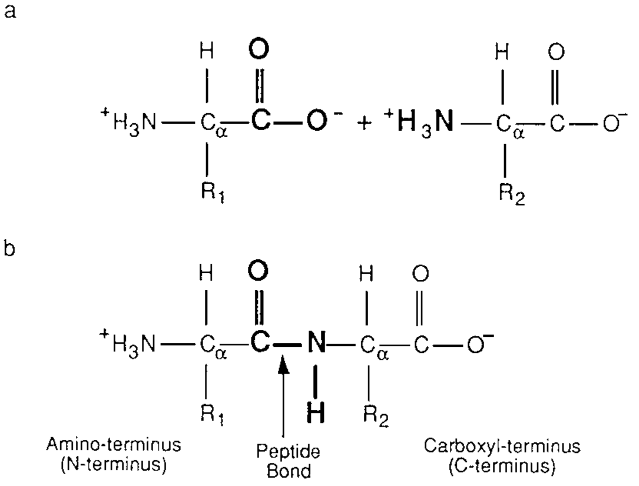

FIGURE 2 Structure of the peptide bond. (a) Two L-amino acids with different side chains, R_1 and R_2; (b) A dipeptide formed from the two amino acids shown in (a). [From Cox, R. A., and Arnstein, H. R. V. (1995). *In* "Encyclopedia of Molecular Biology and Molecular Medicine" (R. A. Meyers, ed.), Volume 6, pp. 108–125. VCH Publishers, New York. With permission.]

involves base-pairing between three bases (designated a codon) in the mRNA and the three-base anticodon of a transfer RNA (tRNA). In a separate reaction, each tRNA is first linked to a particular amino acid and thus the pairing of mRNA with tRNA determines the sequence of amino acids in the resulting protein.

B. Prokaryotic mRNA

In organisms that do not have a nucleus (prokaryotes), pre-mRNA usually undergoes little or no modification so that pre-mRNA and mRNA are very similar if not identical. Because pre-mRNA is colinear with DNA, DNA and proteins are usually colinear in these organisms. Gene expression in prokaryotes usually involves the co-transcription of several adjacent genes, and translation of mRNA sequences into polypeptides may begin at the 5′ end of mRNA while transcription is still in progress at the 3′ end.

C. Eukaryotic mRNA

In cells with a nucleus (eukaryotes) the genetic information is stored mainly in the nucleus and to a minor degree in some organelles (mitochondria and chloroplasts). The description that follows pertains only to nuclear genes. Eukaryotic genes are more complicated than prokaryotic genes because the coding region is often discontinuous: the coding sequences or exons are interrupted by intervening sequences (introns). Thus, genes and proteins are usually *not* colinear in eukaryotes. In the nucleus, a complicated set of splicing reactions removes all the introns from pre-mRNA and fuses the exons into a continuous coding sequence. Other processing steps involve adding a "cap" to the 5′ end of the mRNA and a poly(A) "tail" to the 3′ end. After completion of these nuclear maturation steps the mRNA is transported to the cytoplasm where it is

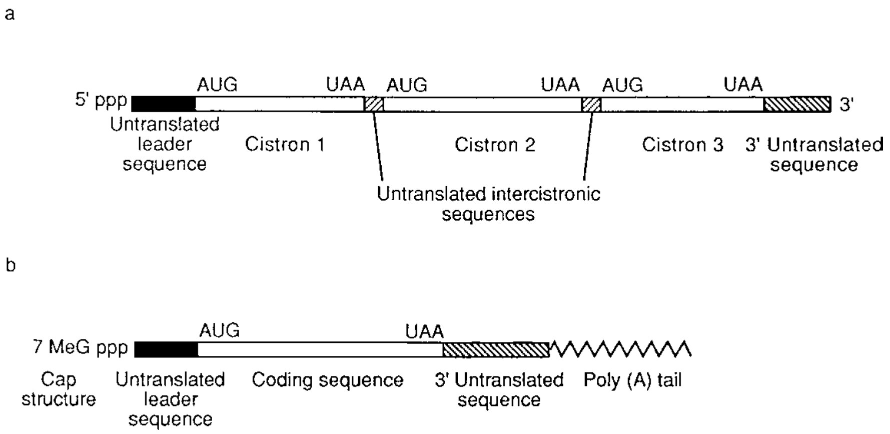

FIGURE 3 Structure of typical mRNAs. (a) *Prokaryotic mRNA.* At the beginning of the transcript there are two additional phosphate residues linked by pyrophosphate bonds to the 5′ phosphate group of the terminal nucleotide (see Fig. 1). The 5′ untranslated sequence often contains a ribosome binding site (the Shine–Dalgarno sequence) which increases the efficiency of translation. AUG and UAA are representative initiation and termination codons, respectively. Cistrons (gene sequences) are separated by short (typically 12 to 24 nucleotides) noncoding intercistronic sequences. (b) *Eukaryotic mRNA.* Typically, the processed transcripts are monocistronic. The 5′ end is usually modified by the addition of 7-MeG, known as a cap structure, which is linked by two pyrophosphate groups to the terminal nucleotide of mRNA. The coding sequence is located between 5′ and 3′ untranslated sequences. In the majority of cases the 3′ end is modified by the addition of 25 to 250 adenylate residues, termed the poly(A) tail. [From Arnstein, H. R. V., and Cox, R. A. (1992). "Protein Biosynthesis," Oxford University Press, Oxford. With permission.]

translated. As with prokaryotic mRNA the coding region is flanked by 5′ and 3′ nontranslated sequences.

D. The Genetic Code

The genetic code is triplet, comma-less, and nonoverlapping. As a consequence, a nucleotide sequence has three possible reading frames (Fig. 4). Because mRNA is normally translated into a unique polypeptide, an essential step in the translation process is the selection of the appropriate reading frame. This is achieved by starting translation at the initiation codon, usually AUG or less frequently GUG, which ensures that the following codons are read in phase within the required reading frame. Of the 64 theoretically possible triplets in the genetic code, 61 sense codons correspond to the 20 genetically encoded amino acids found in all, or nearly all, proteins. When GUG is used as the initiation codon, it codes for methionine by interacting with the anticodon of the initiator Met-tRNA^{Met}, whereas elsewhere it codes for valine. All other codons specify only one amino acid but many amino acids are specified by two or more (up to six) codons; the code is unambiguous, but degenerate. The remaining three codons, termed *nonsense codons*, usually signify termination of synthesis and release of the finished polypeptide chain.

1. Deviations from the Standard Genetic Code

One of the nonsense codons, UGA, has an additional function in the synthesis of selenoproteins. The process involves the initial synthesis of selenocysteyl-tRNA from a novel seryl-tRNA and selenium. This tRNA contains an anticodon that is able to decode UGA and insert selenocysteine residues into the growing polypeptide chain, but only at UGA codons in a particular context of neighboring nucleotides. The insertion of selenocysteine residues into the polypeptide also requires a specific elongation factor T which differs from the factor used for the incorporation of other aminoacyl-tRNAs.

Mitochondria and chloroplasts, as well as certain organisms such as mycoplasma and ciliated protozoa, use a few nonstandard genetic codons. For example, methionine is usually coded for by AUG (or occasionally by GUG) but in human mitochondria this codon is replaced by AUA. Variations in the genetic code are thought to have arisen as a result of the loss of some tRNA genes and mutational pressure on DNA, giving rise to a predominance of either AT- or GC-rich codons.

III. TRANSFER RNA*

By relating individual codons of mRNA to the cognate amino acids, tRNA functions as a key bilingual intermediate in the translation of the genetic code. All transfer RNAs are single-stranded molecules about 80 nucleotides long with a common 3′-terminal CCA sequence. Most of the bases are standard but some (e.g., pseudoU, dihydroU, and T) are derived by modification after transcription of the transfer RNA genes.

The secondary structure of tRNA is usually presented in two dimensions as a cloverleaf to highlight the regions of base-pairing (Fig. 5a). X-ray crystallography reveals that additional hydrogen bonds give rise to an L-shaped tertiary structure (Fig. 5b). The CCA sequence carrying the amino acid is located distal to the anticodon.

The decoding process involves antiparallel base pairing between the three bases of mRNA codons and the complementary anticodons of transfer RNA (tRNA) during peptide bond formation (Fig. 6). The first and middle bases

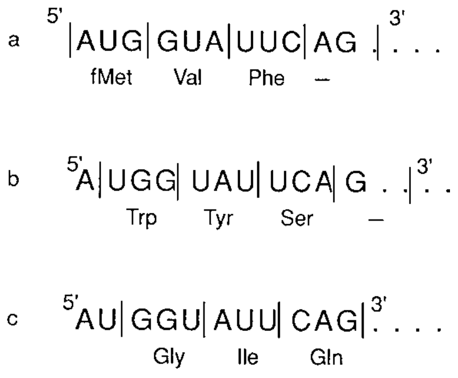

FIGURE 4 Translation of a polynucleotide sequence into three alternative polypeptides using different reading frames. [From Cox, R. A., and Arnstein, H. R. V. (1995). *In* "Encyclopedia of Molecular Biology and Molecular Medicine" (R. A. Meyers, ed.), Volume 6, pp. 108–125. VCH Publishers, New York. With permission.]

*Transfer RNA nomenclature: The amino acid linked to a charged tRNA is indicated by a prefix and the specificity of the transfer RNA (tRNA) in the aminoacylation reaction is shown as a superscript on the right; for example, Phe-tRNA^{Phe} indicates phenylalanine-specific tRNA charged with phenylalanine. The anticodon may be indicated as a right subscript or, alternatively, in the superscript after the amino acid; for example, tRNA$_{UGC}$ or tRNA$^{Ala/UGC}$. The right-hand subscript position is sometimes used to indicate the organism from which the tRNA is derived (e.g., tRNA$^{Val}_{yeast}$). The initiator tRNA, which is specific for methionine, is termed tRNA$^{Met}_{f}$ or tRNA$^{Met}_{i}$. In the cytosol of eukaryotes the charged initiator tRNA is termed Met-tRNA$^{Met}_{f}$ of Met-tRNA$^{Met}_{i}$. Often the superscript ^{Met} is omitted. In prokaryotes and in the mitochondria of eukaryotes, the methionine residue of the charged initiator tRNA is formylated by a transformylase using N^{10}-formyltetrahydrofolate as the donor, giving N-formylmet-tRNA$_f$. Commonly, this charged tRNA is termed fMet-tRNA$_f$. The methionine-specific elongator RNA, which inserts methionine into internal positions of the growing peptide chain, is termed tRNA$^{Met}_{m}$ (or tRNA$_m$) when uncharged, and Met-tRNA$^{Met}_{m}$ (or Met-tRNA$_m$) when charged.

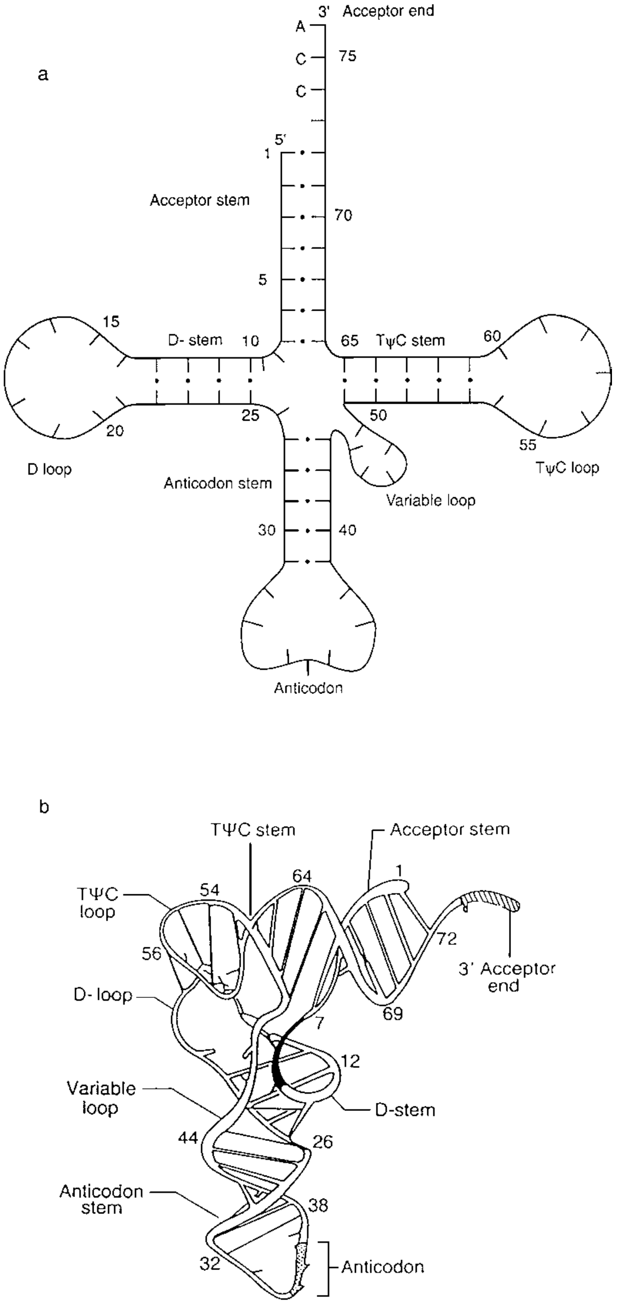

FIGURE 5 Structure of phenylalanyl-transfer RNA. (a) Secondary cloverleaf structure. The 5′ and 3′ ends of the molecule are marked; the continuous line represents the sugar phosphate backbone. The short lines denote base residues and the dots denote base-pairing through standard hydrogen bonds (see Fig. 1). The D loop contains dihydrouracil residues; the TΨC loop contains thymine and pseudouridine. (b) Tertiary structure. The abbreviations for unusual bases are defined in (a). The sugar phosphate backbone is represented by double parallel lines. Standard base pairs are represented by short double lines, and nonstandard base pairs by single lines. The anticodon sequence is stippled, and the acceptor end is shaded. [From Arnstein, H. R. V., and Cox, R. A. (1992). "Protein Biosynthesis," Oxford University Press, Oxford. With permission.]

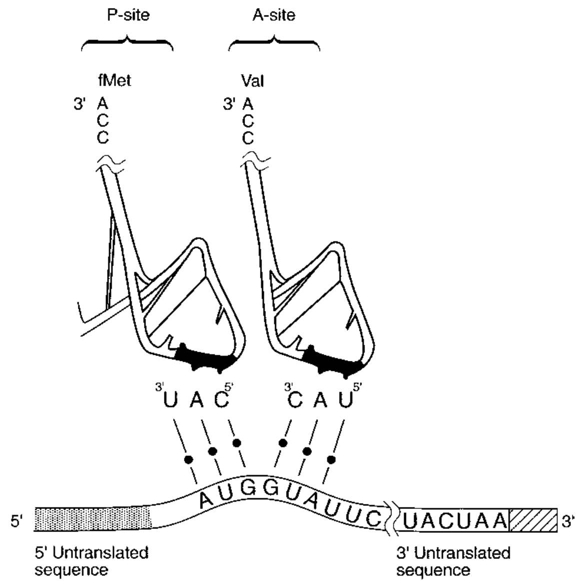

FIGURE 6 Schematic illustration of base pairing between a codon and its anticodon. The diagram shows the interaction at the P site of the ribosome (see Fig. 9) between the initiation codon AUG of mRNA and the anticodon CAU of fMet-tRNA$_f^{Met}$ and the interaction at the A site between the codon GUA of mRNA and the anticodon UAC of Val-tRNA^{Val}. The polarity of an RNA species runs from the 5′ end to the 3′ end. The fragment of tRNA is representative of the general structure (see Fig. 5b) placed in the appropriate orientation; the numbers refer to the nucleotide positions measured from the 5′ end. The interaction between the codon and the anticodon is antiparallel, and the three base pairs have a bihelical conformation.

of the codon form conventional base pairs with the third and middle bases of the anticodon, respectively, but the third base of the codon pairs with the first base of the anticodon by a less stringent interaction (e.g., base-pairing of G with U as well as with C), giving rise to degeneracy. This so-called "wobble" considerably reduces the number of tRNA species required to decode the 61 sense codons. Thus, the protein synthesis system in the cytosol of eukaryotes contains only a few more than 40 different tRNAs, and in mitochondria 22 to 24 tRNA species are sufficient.

The attachment of amino acids to tRNA involves the formation of an ester bond between the alpha-carboxyl group of the amino acid and the 3′-hydroxyl group of the terminal adenosine of tRNA. It requires specific enzymes, the aminoacyl-tRNA synthetases. There are 20 different synthetases, each specific for one of the 20 amino acids, and each enzyme recognizes something unique in the structure of its cognate tRNA. The structural determinants which ensure accuracy of this charging reaction vary for different tRNAs. The anticodon may play a part but sometimes even a single base elsewhere is sufficient to determine the specificity of the tRNA-synthetase interaction. The accuracy of

the synthetase reaction in attaching an amino acid to its cognate tRNA is critically important to the fidelity of the translation process. Once the aminoacyl-tRNA has been formed, the subsequent incorporation of the amino acid residue into a polypeptide does not depend on the amino acid itself but only on the interaction between the anticodon of the aminoacyl-tRNA with the codon of mRNA. Thus, an error in the synthetase reaction would lead to the incorporation of an inappropriate amino acid into the polypeptide.

Synthesis of aminoacyl-tRNA (III) from amino acids (I) requires activation of the amino acid carboxyl group with formation of an intermediate enzyme-bound aminoacyladenylate (II).

$$\mathrm{E} + \underset{\text{(I)}}{\mathrm{R{-}\overset{\overset{\displaystyle H}{|}}{\underset{\underset{\displaystyle NH_3^+}{|}}{C}H}{-}CO_2^-}} + \mathrm{ATP} \rightarrow \underset{\text{(II)}}{\mathrm{E\cdot R{-}\overset{\overset{\displaystyle H}{|}}{\underset{\underset{\displaystyle NH_3^+}{|}}{C}H}{-}CO\cdot AMP}} + \underset{\underset{\displaystyle 2P_i}{\downarrow}}{\mathrm{PP_i}}$$

$$\rightarrow \underset{\text{(III)}}{\mathrm{R{-}\overset{\overset{\displaystyle H}{|}}{\underset{\underset{\displaystyle NH_3^+}{|}}{C}H}{-}CO\cdot tRNA}} + \mathrm{AMP} + \mathrm{E}$$

where E = aminoacyl-tRNA synthetase.

The energy for the reaction (two high-energy phosphate bonds) is provided by ATP and stored in the ester bond of the aminoacyl-tRNA to be used subsequently for peptide bond synthesis.

IV. RIBOSOME STRUCTURE AND FUNCTION IN TRANSLATION

A. Ribosome Structure

Ribosomes are high-molecular-weight complexes of RNA (rRNA) and proteins (Table I), and the electron-dense particles are easily visualized by electron microscopy. Ribosomes from various sources (prokaryotes, eukaryotic cytoplasm, mitochondria, chloroplasts, and kinetoplasts) vary in size from 20 to 30 nm in diameter, but all are composed of a large and a small subparticle or subunit and perform similar functions in protein synthesis. The principal functional domains of the ribosome and associated components are given in Fig. 10. More detailed resolution of the ribosome structure has allowed the placement of mRNA, aminoacyl-tRNA, peptidyl-tRNA, and the nascent polypeptide chain (see Section C and Fig. 11).

The small subunit comprises a single rRNA of 0.3–0.7 $\times$ 10^6 Da and single copies of 20 to 30 unique proteins. It has a major function in binding initiator tRNA and mRNA in the initiation of protein synthesis and in decoding the genetic message.

The large subunit comprises a high-molecular-weight rRNA (0.6–1.7 $\times$ 10^6 Da) and often one or two smaller rRNAs (0.03–0.05 $\times$ 10^6 Da) and 30 to 50 different proteins are present, with one exception, as single copies. The large subunit binds aminoacyl-tRNA at the A-site, peptidyl-tRNA at the P site, and discharged tRNA at the E (exit) site. The large subunit contains the peptidyl transferase and, unusually, this enzyme activity resides in the rRNA molecule itself rather than in the associated ribosomal proteins. This subunit is also involved in binding elongation factor G which is required for translocation (Fig. 10).

B. The Ribosome Cycle in Translation

Polypeptide synthesis can be divided into three stages: initiation, elongation, and termination (see Fig. 7). Initiation involves the binding of a ribosome to mRNA with the initiation codon correctly placed in the P-site. Elongation leads to the stepwise increase in the length of the polypeptide chain through the transfer of the growing chain to the amino group of aminoacyl-tRNA. Termination of chain elongation and release of the completed polypeptide occurs when a termination codon reaches the A site. All stages require the participation of protein factors. Advances in establishing the structures of translational factors by X-ray crystallography, which have been rapid in the last decade, are reviewed by Al-Karadaghi *et al.* (2000).

1. Formation of Pre-Initiation Complexes

The ribosome cycle starts with the stepwise formation of an initiation complex from mRNA, charged initiator tRNA, and ribosomal subunits. A number of pre-initiation complexes are formed as intermediates and the process is facilitated by initiation factors. In outline, prokaryotic and eukaryotic systems are similar, but there are a few differences, particularly as regards the complexity of the initiation factors and details of the mechanisms.

a. Prokaryotic systems. Three proteins, initiation factors IF-1, IF-2, and IF-3 (see Table II), are required for the initiation of protein biosynthesis (see Fig. 8a). Ribosomal subunits are released by dissociation of ribosomes following translation of the mRNA. Dissociation is facilitated by the combined action of the initiation factors IF-1 and IF-3; IF-1 increases the rate of dissociation and

TABLE I Properties of Ribosomes and Ribosomal Subunits

Source	$S_{20,w}$	Size (nm)[a]	Mass (Mda)	RNA:protein (w/w)	Axial ratio	rRNA mass (Mda)	$S_{20,w}$ (nominal)
			Ribosomes				
Prokaryotes							
E coli	70	22.5 ± 2.5	2.6–2.9	2:1	—[f]		
Eukaryotes							
Cytoplasm	80	28.0 ± 2.8	3.4–4.5	1:1	—		
Chloroplast	70	22.5 ± 2.5	2.5–3.3	1:1	—		
Mitochondria[b]							
Protists[c]	60–80		*ca.* 3.25	1:1.38	—		
Animals	55–60		2.7–3.2	1:1.7–1:4	—		
Fungi	67–80		4.2	1:1.3–1:1.6	—		
Higher plants	78		—	—	—		
			Small ribosomal subunits				
Prokaryotes							
E. coli	30	22.0 ± 2.2	0.95	2:1	2:1	0.6	16
Eukaryotes							
Cytoplasm	36–41	25.0 ± 2.5	1.4	1:1	2:1	0.7	18
Chloroplast	28–35	22.0 ± 2.2	1.2	1:1	2:1	0.6	16
Mitochondria							
Protists	32–45		—	—	—	0.2–0.35	11–16
Animals	32–40		1.1–1.6	—	—	0.31	12–14
Fungi	32–29		1.6–1.7	—	—	0.45–0.64	15–19
Higher plants	44		—	—	—	0.64	18
			Large ribosomal subunits				
Prokaryotes							
E. coli	50	22.5 ± 2.2	1.75	2:1	1:1	1.17	23
						0.04	5
Eukaryotes							
Cytoplasm	60	28.0 ± 2.8	2.1–3.1	1:1	1:1	1.2–1.75	25–28
						0.05	5.8
						0.4	5
Chloroplasts	46–54	22.5 ± 2.2	2.4	1:1	1:1	1.1	23
						0.04	5
						0.03[d]	4.55[d]
Mitochondria							
Protists	45–60		—	—	—	0.4–0.74	12–24
Animals	25–35		1.65–2.0	—	—	0.5	16–21
Fungi	50		—	—	—	0.77–1.22	21–25
Higher plants	60		—	—	—	1.25	26
						0.04[e]	5[e]

[a] Largest dimension.

[b] Mitochondrial ribosomes are preferentially associated with the inner mitochondrial membrane and are more diverse than cytoplasmic or chloroplast ribosomes. Particular features include post-transcriptional oligoadenylation of the 3′ ends of both the smaller and larger rRNA components. Except for mitochondrial ribosomes of higher plants, the essential sequence motifs 5S rRNA and 5.8S rRNA are present in the large rRNA component and are not found as individual species.

[c] Mitochondria of protists such as *Trypanosoma brucei* are also known as kinetoplasts.

[d] 4.5S rRNA corresponding to the 100 nucleotides at the 3′ end of the 23S rRNA in eubacteria occurs as a separate species in chloroplasts of higher plants.

[e] This species is present only in plant mitochondria and is absent from all other mitochondrial ribosomes.

[f] Data not available.

TABLE II Prokaryotic Initiation Factors from *E. coli*

Factor	M_r (kDa)	Properties and function
IF-1	9	Stimulates activity of IF-2; accelerates dissociation of unprogrammed ribosomes to subunits.
IF-2	100	Binds fMet–tRNA$_f$ to the ribosomal P site by a GTP-requiring reaction.
IF-3	22	Binds natural mRNAs to the small ribosomal subunit probably by facilitating base-pairing between the untranslated leader sequence and the 3′ end of 16S rRNA; prevents ribosomal subunit association when bound to the small subunit.

From Arnstein, H. R. V., and Cox, R. A. (1992). "Protein Biosynthesis," Oxford University Press, London. With permission.

IF-3 acts as an anti-association factor when bound to the 30S ribosomal subunit, thereby displacing the equilibrium in favor of subunit formation. Initiation factor IF-2 is also able to bind to the 30S subunit and this association is stabilized by IF-1 and GTP, the latter acting as a steric effector without being hydrolyzed at this stage. IF-2 plays a central role in binding fMet-tRNA$_f$ to the 30S pre-initiation complex by specific recognition of the N-formylmethionine residue attached to the initiator tRNA, thus restricting this interaction to charged initiator tRNA. All three factors bind to the 30S ribosomal subunit near the 3′ end of the 16S ribosomal RNA at adjacent sites that are located at the interface between the small and large ribosomal subunits.

In the next step, the initiator tRNA and mRNA associate with the 30S–IF-1–IF-2–IF-3 complex with release of IF-3. There is evidence from *in vitro* experiments that the binding of mRNA precedes that of the initiator tRNA.

Messenger RNA binds to the small ribosomal subunit immediately before formation of the final initiation complex with the initiation codon correctly positioned in the P-site (see Fig. 7a). In the case of bacterial and bacteriophage messengers, the molecular recognition mechanism proposed by Shine and Dalgarno (1974) involves basepairing between short nucleotide sequences, most often CUCC, near the 3′ end of the 16S ribosomal RNA and a complementary region, usually consisting of 3 to 9 bases on the 5′ side of the mRNA initiation codon, which has been found to be present in nearly all of more than 150 bacterial and bacteriophage messengers. Studies with mutants and mRNA fragments indicate that, in addition to the Shine–Dalgarno interaction, outlying upstream sequences in the leader region may also provide recognition signals between mRNAs and ribosomes, possibly by ensuring that the Shine–Dalgarno sequence is in an appropriate conformation.

The Shine–Dalgarno mechanism is also found in chloroplast protein synthesis as judged from sequence analysis of the 16S rRNA and mRNAs, but apparently not in the mammalian mitochondrial system where the initiator codon occurs either directly at, or only a few nucleotides downstream from, the 5′ end of mRNA, which excludes the possibility of mRNA–rRNA base-pairing in this region.

b. Eukaryotic systems. At least 12 proteins, the eukaryotic initiation factors (eIF) (see Table III), are needed for initiation of protein biosynthesis (Fig. 8b). The dissociation of cytosolic 80S ribosomes is facilitated by a complex initiation factor, eIF-3 (M_r approx. 5–700,000), consisting of 9 to 11 polypeptide chains, which binds to the small ribosomal subunit (40S) and prevents its re-association to 80S ribosomes. Thus, this factor has anti-association activity, but low-molecular-weight proteins with similar activity have also been reported, and a protein, eIF-4C, of M_r 20,000, seems to function as an accessory factor to eIF-3 in the formation of a 43S ribosomal pre-initiation complex. Also, another protein factor, eIF-6, of M_r 24,000, prevents re-association by binding to the large (60S) ribosomal subunit.

Initiation factor eIF-2 gives a stable binary complex with GTP which binds the initiator tRNA, Met-tRNA$_f$, forming a ternary complex. Interaction of this ternary complex with the 40S ribosomal subunit containing bound initiation factors eIF-3 and eIF-4C gives rise to the 43S pre-initiation complex, which is competent to bind messenger RNA in the presence of three further initiation factors, eIF-4A, eIF-4B, and eIF-4F, together with ATP.

The binding of cytosolic eukaryotic messenger RNAs to the small ribosomal subunit probably does not involve base-pairing with the 18S rRNA, as no uninterrupted sequences of the Shine–Dalgarno type have been found. Instead, a "scanning model" has been proposed, in which the pre-initiation complex, composed of the 40S ribosomal subunit, Met-tRNA$_f^{Met}$ and associated initiation factors, binds at or near the 5′ cap of the mRNA and slides along the messenger until it encounters the first AUG triplet, at which point the 60S ribosomal subunit joins to give rise to the 80S initiation complex. Recognition of the cap is facilitated by cap-binding proteins, which mediate an ATP-dependent melting of the mRNA secondary structure at the 5′-terminal region to allow the mRNA to thread through a channel in the neck of the 40S subunit. The cap structure is required for efficient binding and translation even in cases where the initiating AUG codon occurs hundreds of nucleotides downstream. As a rule, scanning by the 40S subunit stalls at the first AUG codon, which is recognized mainly by interaction with the anticodon

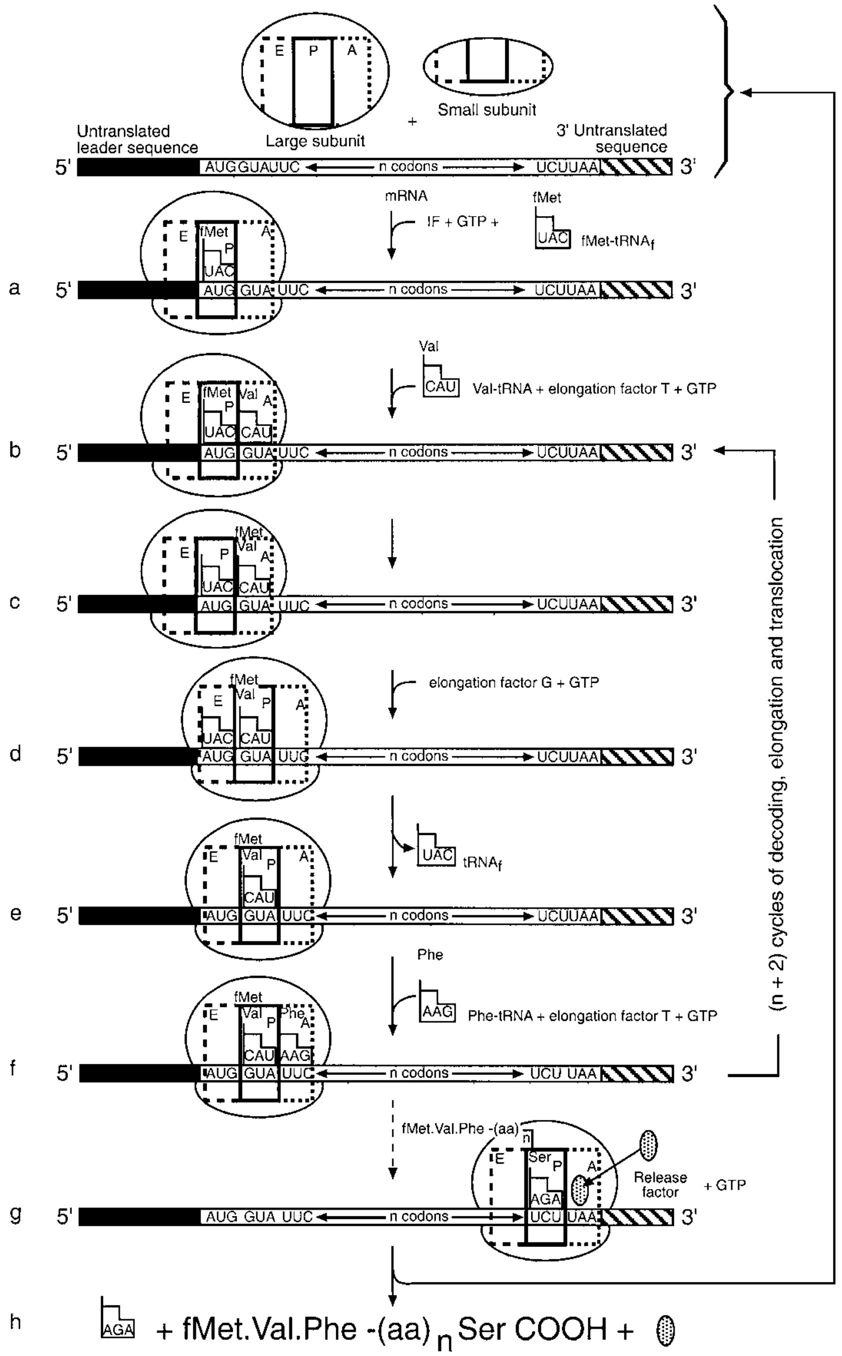

FIGURE 7 Schematic view of the biosynthesis of a polypeptide by translation of prokaryotic mRNA. (a) Formation of the initiation complex from mRNA, ribosomal subunits, initiation factors (IF), GTP, and fMet-tRNA^{Met} (for details, see Fig. 8). The binding sites for aminoacyl-tRNA, peptidyl-tRNA, and uncharged tRNA are designated A, P, and E, respectively. (b) Decoding of the codon GUA with Val-tRNA^{Val}. (c) Formation of the peptide bond by transfer of fMet to Val-tRNA^{Val} forming fMet–Val-tRNA^{Val}. (d) Translocation of tRNA$_f^{Met}$ to the E site and of fMet–Val-tRNA^{Val} to the P site with codon UUC aligned with the A site. (e) Ejection of tRNA$_f^{Met}$ from the ribosome. (f) Decoding of codon UUC with Phe-tRNA^{Phe}. (g) Decoding of codon UAA with release factor. Steps (b) to (f) constitute the elongation–translocation cycle. The position shown is reached by repeating the cycle $n + 2$ times after formation of fMet. Val-tRNA. (h) Release of the completed polypeptide, fMet–Val–Phe–(aa)$_n$–Ser, ribosomal subunits, release factor, mRNA, and tRNA^{Ser}. The cycle may then be repeated starting at (a). [From Cox, R. A., and Arnstein, H. R. V. (1997). *In* "Encyclopedia of Molecular Biology and Molecular Medicine," (R. A. Meyers, ed.), Volume 6, pp. 108–125. VCH Publishers, New York. With permission.]

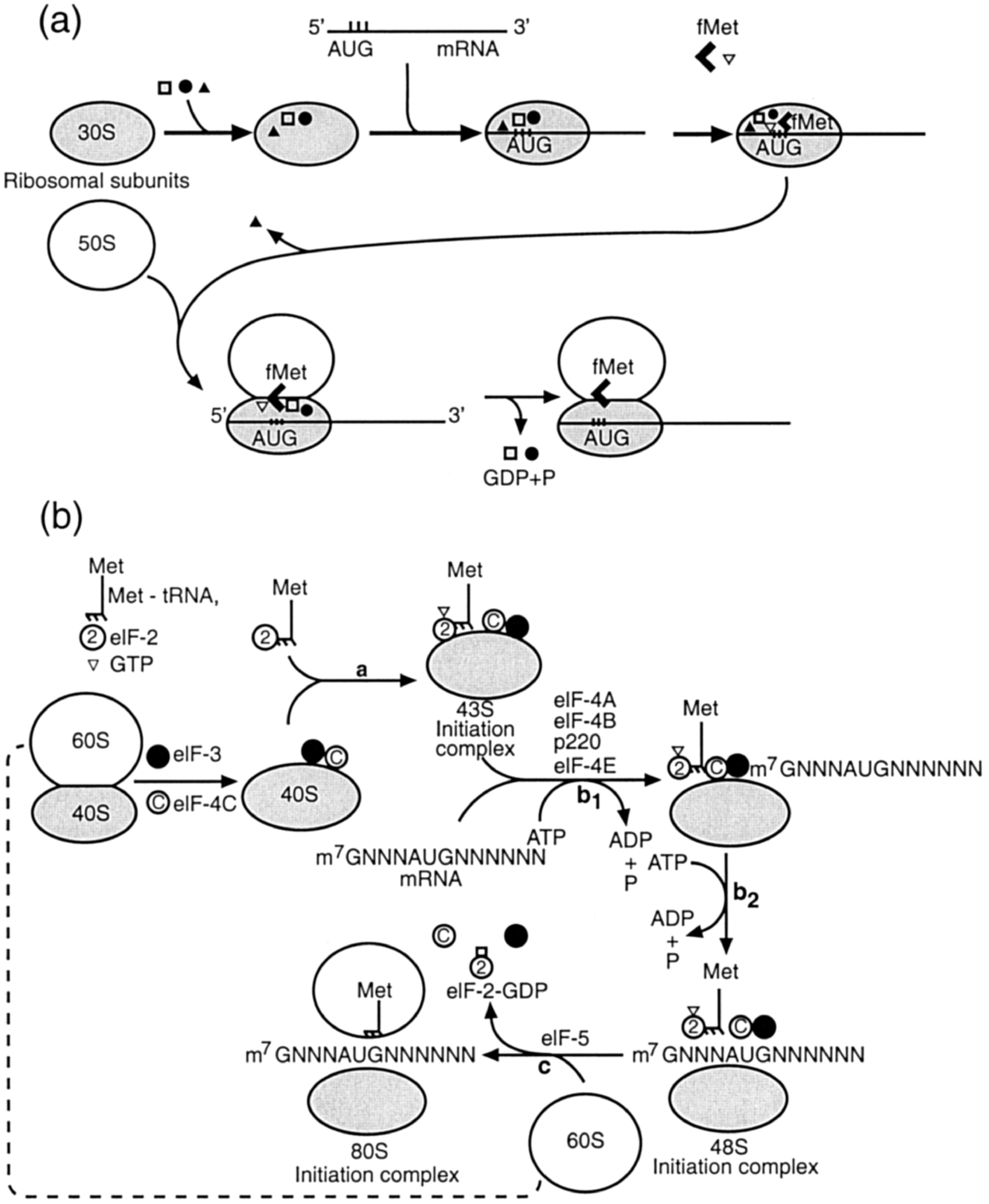

FIGURE 8 Schematic diagram illustrating the formation of initiation complexes. (a) Prokaryotic initiation. The following symbols are used for initiation factors and other components involved in the formation of the 70S initiation complex: ■, IF-1; ●, IF-2; ▲, IF-3; ❮, fMet-tRNA$_f$, ▽, GTP; other symbols are defined in the figure. (b) Eukaryotic initiation complexes in mammalian protein biosynthesis from components defined in the figure. The process may be divided into three stages: **a**, formation of a 43S pre-initiation complex; **b**, binding of mRNA with formation of a 48S pre-initiation complex; and **c**, synthesis of the 80S initiation complex containing the initiator tRNA in the correct position for peptide bond formation. [From Arnstein, H. R. V., and Cox, R. A. (1992). "Protein Biosynthesis," Oxford University Press, Oxford. With permission.]

of the Met-tRNA$_f^{Met}$. However, this recognition also depends in some way on eIF-2 and may be modulated as a result of deviation of the mRNA structure from the consensus sequence GCCGCC^A/$_G$CCAUGG. Sequence context may also account for the rare cases where initiation occurs downstream of the 5′-proximal AUG codon. Where this sequence context is unfavorable, initiation becomes inefficient, hence most 40S subunits will tend to initiate further along the mRNA at another AUG triplet in a more favorable context. This model also explains rare cases where initiation is not restricted to one particular AUG codon and translation of a single mRNA gives rise to two proteins.

TABLE III Eukaryotic Initiation Factors

Initiation factor	Synonym	M_r (kDa) of factors and subunits	Properties and function
eIF-1		15	Stabilizes initiation complexes
eIF-2		alpha, 38[a] beta, 35–50[a] gamma, 55	GTP-dependent binding of Met-tRNA$_f$ to the small ribosomal subunit
eIF-2B	GEF	27, 37, 52, 67, 85[a]	Conversion of eIF-2–GDP into eIF-2–GTP
eIF-3		9–11 subunits 24–170[a]	Associates with 40S subunit to maintain dissociation; binds mRNA to 43S preinitiation complex
eIF-4A	50-kDa component of CBP-II	50	ATP-dependent unwinding of the secondary structure of the mRNA 5′ region; stimulates translation of exogenous mRNA in cell-free systems
eIF-4B		80[a]	mRNA binding; stimulates cell-free translation; ATPase activity of eIF-4A and eIF-4F; AUG recognition and recycling of eIF-4F
eIF-4C		17	Ribosome dissociation; 60S subunit joining
eIF-4D		17	Formation of first peptide bond
eIF-4E	CBP-I; 24-kDa CBP	24–28[a]	Binds mRNA cap structure
eIF-4F	CBP-II; cap-binding protein complex	24 (CBP-I),[a] 50 (eIF-4A), 200[a]	ATPase; unwinds mRNA secondary structure; stimulates cell-free translation
eIF-4G		220	Stimulates protein synthesis by interacting with eIF-4E and poly(A) protein to circularize polysomes
eIF-5		60[a]	GTPase; release of eIF-2 and eIF-3 from pre-initiation complex to allow joining of the 60S subunit
eIF-6		24	Anti-association activity; binds to the 60S ribosomal subunit

[a] Denotes subunit can be phosphorylated *in vivo*.

Abbreviations: GEF, guanine nucleotide exchange factor; CBP, cap-binding protein.

2. Initiation Complex Formation: Joining of the Large Ribosomal Subunit

The last event in the initiation of protein synthesis involves the joining of the large ribosomal subunit to the pre-initiation complex (Fig. 8). In the prokaryotic system, association of the 50S subunit with the 30S pre-initiation complex takes place with hydrolysis of GTP by the GTPase activity of IF-2 and release of IF-1, IF-2, GDP, and P_i. GTP hydrolysis is essential for the release of IF-2 from the initiation complex, which is a prerequisite for allowing the fMet-tRNA$_f$ to engage in the formation of the first peptide bond. In eukaryotic protein synthesis, the 80S initiation complex is formed by joining the 60S ribosomal subunit to the 48S pre-initiation complex consisting of the 40S ribosomal subunit, eIF-2, eIF-3, GTP, Met-tRNA$_f$, mRNA, and possibly eIF-4C. This coupling reaction requires an additional factor, eIF-5, which mediates the hydrolysis of GTP to GDP with release of eIF-2-GDP, P_i, and eIF-3 from the 48S pre-initiation complex.

By this stage, all initiation factors have been released and are available for recycling, although the exact steps at which factors are released from intermediate complexes are not known in every case. There is thus an initiation factor cycle within the ribosome cycle, and regulation of the activity of factors, particularly eIF-2, is an important control mechanism in translation (see Section V.D.1).

In the initiation complex, location of the charged initiator tRNA in the P site of the ribosome allows transfer of the methionine residue to the amino group of another aminoacyl-tRNA in the A site (Fig. 7b) by peptidyl transferase to form dipeptidyl-tRNA (see Fig. 7c). Functional insertion of Met-tRNA$_f$ directly into the P site can be demonstrated using the trinucleotide AUG as a synthetic mRNA and another trinucleotide, for example UUU, to bind an acceptor aminoacyl-tRNA (in this case Phe-tRNA).

It is possible to measure the peptidyl transferase activity of the large subunit in the absence of mRNA by using the antibiotic puromycin, which resembles the 3′-terminal region of Phe-tRNA in structure, as an artificial acceptor to form methionyl puromycin from Met-tRNA$_f$.

3. Polypeptide Chain Synthesis: The Elongation–Translocation Cycle

This cycle is outlined in Figs. 7b–f.

a. Elongation. The first peptide bond is formed when the aminoacyl-tRNA in the ribosomal A site is converted into the corresponding methionyl-aminoacyl-tRNA by transfer of the methionyl (or N-formylmethionyl) residue from the charged initiator tRNA in the P site (Fig. 7c). In

artificial cell-free systems, any N-substituted aminoacyl-tRNA, such as peptidyl-tRNA or N-acetylaminoacyl-tRNA, can function in peptide bond synthesis as a donor in the P site in place of the charged initiator tRNA. The reaction is catalyzed by the peptidyltransferase activity of the large ribosomal subunit. No soluble cofactors appear to be involved, but monovalent cations (K^+) at a concentration of 100 m*M* or more and divalent cations (Mg^{2+}) below 2 m*M* are required.

Efficient entry of aminoacyl-tRNA into the ribosomal A site requires the participation of an elongation factor, termed EF-Tu in prokaryotes (see Table IV) and EF-1 (EF-1_L) in eukaryotes (see Table V), and GTP. This elongation factor forms a ternary complex with GTP and all aminoacyl-tRNAs except initiator tRNA, but not with uncharged tRNA, thus ensuring that only appropriately charged tRNAs are efficiently bound in the A site. A special elongation factor showing extensive homology with both EF-Tu and IF-2 is involved in the synthesis of selenoproteins (see Section II.C) from selenocysteyl-$tRNA^{UCA}$ in *E. coli*.

The above-mentioned aminoacyl-tRNA binding reaction catalyzed by EF-Tu is the rate-limiting step in the elongation cycle; peptide bond formation and translocation are much faster. The initial binding of the ternary complex to the ribosome is readily reversed, but the interaction is stabilized by the subsequent codon recognition which induces the GTPase conformation of EF-Tu leading immediately to the hydrolysis of the GTP component of the ternary complex to GDP. Hydrolysis of the GTP moiety causes a further change in the conformation of EF-Tu from the GTP-binding to the GDP-binding form. This conformational change leads to the release of aminoacyl-tRNA, allowing its CCA end to align with the peptidyl transferase center of the ribosome and the instantaneous formation of the peptide bond. The elongation factor is later released from the ribosome as a complex with GDP. The operation of EF-Tu is thus similar to that of initiation factor 2 which binds charged initiator tRNA to the small ribosomal subunit.

TABLE IV Properties of Prokaryotic Elongation and Termination Factors from *E. coli*

	M_r (kDa)	Properties and function
Elongation factors		
EF-T_u	43	N-terminal acetyl-serine; heat labile; binds aminaocyl-tRNA to the ribosomal A site
EF-T_s	30	Heat stable; regeneration of EFT_u–GTP
EF-G	77	GTP-dependent translocation of peptidyl-tRNA and its mRNA codon from the A site to the P site of the ribosome
Termination (release) factors		
RF1	36	Requires UAA or UAG codons for hydrolysis of peptidyl-tRNA
RF2	38	Requires UAA or UGA codons for hydrolysis of peptidyl-tRNA
RF3	46	Enhances RF1 and RF2 activity

TABLE V Eukaryotic Elongation and Termination Factors

	M_r (kDa)	Properties and function
Elongation factors from various yeast, animal, and plant cells		
eEF-1A (EF-1_L or eEF-T_u)	50–60	Analogous to EF-T_u
eEF-1B (eEF-T_s)	30	Analogous to EF-T_s
eEF-2	105	Contains essential SH groups and one residue of a post-translationally modified histidine residue; GTP-dependent translocation analogous to EF-G
eEF-3	125	GTPase and ATPase activity; function not fully defined
Termination or release factors		
eRF-1	110	Two 55-kDa subunits; binds to the ribosome A site by a GTP and termination codon dependent reaction; hydrolyzes peptidyl-tRNA in the P site
eRF-2		GTPase; stimulates eRF-1 activity

Following dissociation from the ribosome the EF-Tu–GDP complex interacts with another elongation factor, EF-Ts, with formation of an EF-Tu–EF-Ts heterodimer and release of GDP. Reaction of the heterodimer with GTP regenerates the EF-Tu–GTP complex required for binding aminoacyl-tRNA. The sequence of events is similar in eukaryotes with eEF-1A (Mr 50,000) corresponding to EF-Tu and eEF-1B (Mr 30,000) to EF-Ts.

Selection of the specific aminoacyl-tRNA to be bound at the ribosomal A site is by base-pairing between the relevant mRNA codon and the tRNA anticodon. Because this interaction involves only a triplet of bases and hence a maximum of nine hydrogen bonds (see Fig. 1B), it is intrinsically unstable at physiological temperatures and is probably stabilized by components of the ribosome to allow sufficient time for peptide bond synthesis to occur. Also, the codon–anticodon pairing must be monitored for fidelity in order to minimize errors in translation. In *E. coli* there is genetic and biochemical evidence that one of the proteins of the small ribosomal subunit, S12, is involved in ensuring the fidelity of normal translation and in causing the mistranslation which occurs in the presence of the antibiotic streptomycin due to incorrect codon-anticodon interactions.

b. Translocation. Translocation involves the movement of the ribosome along the mRNA in the $5' \rightarrow 3'$

direction. Immediately after synthesis of the first peptide bond, the ribosomal A site contains dipeptidyl-tRNA while uncharged initiator tRNA remains in the P site. Thus, both these sites are occupied, and to allow the next aminoacyl-tRNA to enter the A site it is necessary to eject the uncharged tRNA and shift the dipeptidyl-tRNA from the A into the P site. This translocation (Fig. 7d) takes place as a concerted process involving movement of both messenger RNA and dipeptidyl RNA together into the P site, leaving the A site occupied by the next mRNA codon and free to accept the cognate aminoacyl-tRNA (see Fig. 7e). At the same time, the deacylated tRNA moves first into an E (exit) site with subsequent ejection when the next aminoacyl-tRNA enters the A site.

Translocation requires the participation of another elongation factor (EF-G in prokaryotes and EF-2 in eukaryotes) and GTP (Table IV). It seems that when EF-G and GTP bind to the ribosome, translocation occurs but GTP hydrolysis is required only subsequently to release EF-G and GDP. The location of the EF-G binding site on the ribosome overlaps with that for EF-Tu, thus EF-G must be released before the EF-Tu–aminoacyl-tRNA–GTP complex can enter the A site. Analogous reactions occur in eukaryotic systems.

There is little information about the details of the translocation mechanism. A continuous polyribonucleotide chain is not essential, as translocation can occur with individual trinucleotides. It seems likely that movement of the mRNA is dependent on and tightly coupled to that of the tRNA with the binding sites for the tRNA providing the precision for movement by exactly one codon. Presumably, binding of EF-G and GTP after release of EF-Tu–GDP following peptide bond synthesis induces a conformational change in the ribosome which leads to translocation.

After translocation the ribosomal P site is occupied by dipeptidyl-tRNA and the vacant A site contains the third mRNA codon. Entry of the next aminoacyl-tRNA, selected as before by the codon–anticodon interaction, into the A site (Fig. 7f) enables peptide bond synthesis to continue and repeated operation of the elongation-translocation cycle gives rise to a stepwise elongation of the nascent polypeptide chain, each complete cycle elongating the chain by one amino acid residue and moving the mRNA by one codon in the 5′ to 3′ direction. When the end of the coding sequence is reached and one of the termination (or stop) codons has entered the A site, translation stops and the completed polypeptide chain is released.

c. Termination. (See Figs. 7g–h.) The presence of one of the three termination codons, UAA, UAG, or UGA, in the A site results in the binding of a release factor (Table IV) instead of an aminoacyl-tRNA to the ribosome. In prokaryotes, two release factors have been identified, one (RF1) recognizing UAA and UAG, the other (RF2) functioning with UGA. Ribosomal binding and release of RF1 and RF2 are stimulated by a third factor, RF3, which interacts with GTP and GDP. In eukaryotic cells such as reticulocytes, one release factor (eRF) has been found to function with all three termination codons, and the binding of this factor to ribosomes is stimulated by GTP but not GDP. Although the details are not entirely clear, GTP hydrolysis appears to be required for the release of the finished polypeptide chain by cleavage of the peptidyl-tRNA bond and completion of the termination process leading to dissociation of the release factor from the ribosome.

Thus, at the end of the ribosome cycle the coding sequence of messenger RNA has been translated to produce a particular polypeptide chain, and all the components involved become available for re-use in another round of the cycle (Fig. 7h). Usually, several ribosomes become attached to one mRNA molecule, giving rise to polyribosomes (also called *polysomes*; Fig. 9). In eukaryotic cells the efficiency of protein synthesis is stimulated by factor eIF4-G (Table III), which interacts with both factor eIF-4E and a poly A-binding protein. The resultant circularized polysomes show an enhanced ability to re-initiate after release of the ribosomal subunits from the messenger RNA at the end of a round of translation.

C. High-Resolution Structural Studies of the Ribosome

Electron microscopy of ribosomes has provided sufficient information to allow the construction of models showing the general features of the principal functional domains, such as the location of mRNA, ribosomal subunits, and factors (Fig. 10).

During the last ten years, X-ray diffraction studies have led to considerable advances in the elucidation of the structure of the ribosome in more detail, culminating in the determination of the *E. coli* ribosome at 0.78-nm resolution (Fig. 11) and of the small and large subunits at resolutions of 0.3 nm and 0.25 nm, respectively. The structures have revealed the identity of each amino acid and each nucleotide. The findings provide insights at the atomic level into the reactions leading to the decoding of mRNA and the formation of the peptide bond. Moreover, the importance of the role of rRNA in ribosome function has become evident from the structures; the functional regions of both small and large subunits are rich in RNA. The three-dimensional structures also highlight the dynamic aspects of ribosome function leading to the view that the ribosome is a highly sophisticated motor driven by GTP with rRNA playing a leading role.

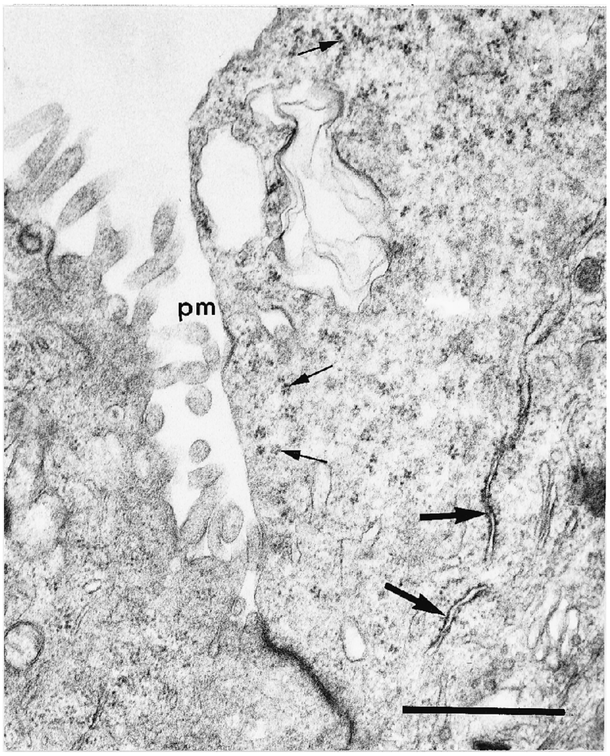

FIGURE 9 Electron micrograph of a thin section through two neighboring epithelial cells showing endoplasmic reticulum (large arrows) to which are attached numerous polysomes. Groups of free ribosomes (small arrows) occur in the cytoplasm; pm, plasma membrane. Bar represents 1000 nm. [Micrograph by I. D. J. Burdett, National Institute for Medical Research, London, U.K.]

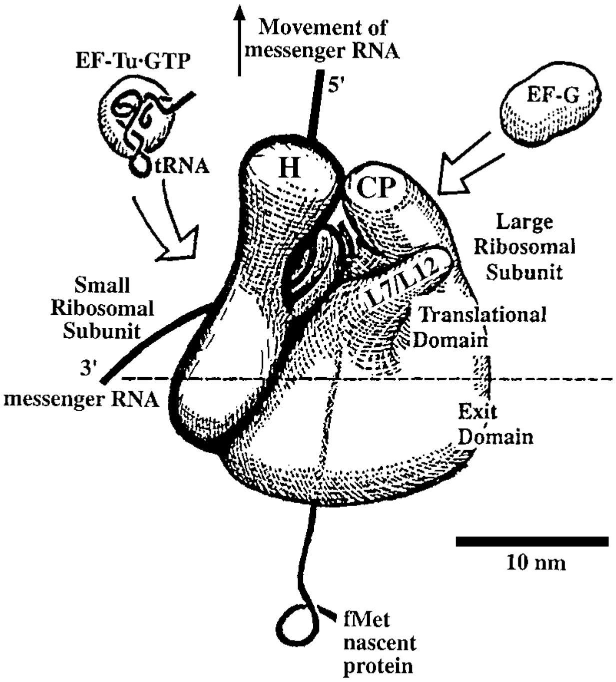

FIGURE 10 Model of the *E. coli* ribosome, based on low-resolution electron microscopy. The diagram shows the relative orientation of the large and small subunits and other functional components involved in polypeptide chain synthesis. H, head of the small subunit; CP, central protuberance and L7/L12 stalk of the large subunit; EF-Tu, elongation factor required for binding aminoacyl-tRNA to the ribosomal A site; EF-G, elongation factor required for translocation of the peptidyl-tRNA from the A to the P site. The broken line indicates the boundary between the translational and exit domains which are involved in peptide bond formation and extrusion of the nascent polypeptide chain, respectively. [From Cox, R. A., and Arnstein, H. R. V. (1997). *In* "Encyclopedia of Molecular Biology and Molecular Medicine" (R. A. Meyers, ed.), Volume 6, pp. 108–125. VCH Publishers, New York. With permission.]

The higher resolution studies of subunits have also revealed the mode of action of several antibiotics such as paromomycin, streptomycin, and spectinomycin, which modify the decoding function of the small subunit, and chloramphenicol, puromycin, and vernamycin, which affect peptide bond formation.

V. TRANSLATIONAL CONTROL OF GENE EXPRESSION

Cells needs to synthesize specific proteins in the required amounts at particular times and to deliver them to the correct locations. These processes depend on a great many interactions and numerous mechanisms exist for the translational control of gene expression. Examples of the ways used by cells to control the translation of mRNA are presented in the sections that follow. This material is intended to be illustrative rather than comprehensive, and it is to be expected that novel control mechanisms will continue to be discovered.

A. mRNA Stability

Provided all other components of the translational systems are present in optimum amounts, control of translation may be achieved through the availability of the relevant messenger RNAs at the site of protein synthesis. The steady-state level of mRNA is determined by the rate of its synthesis and degradation. Control of transcription is of major importance for synthesis in both prokaryotes and eukaryotes. The stability of mRNA depends on its primary and secondary structure as well as on the presence of factors such as stabilizing proteins and nucleases.

The secondary structure of mRNA is determined by its nucleotide sequence. For any mRNA a number of coding sequences are possible because of the degeneracy of the genetic code. Thus, degeneracy allows for particular features of secondary structure (often termed *cis* factors) to

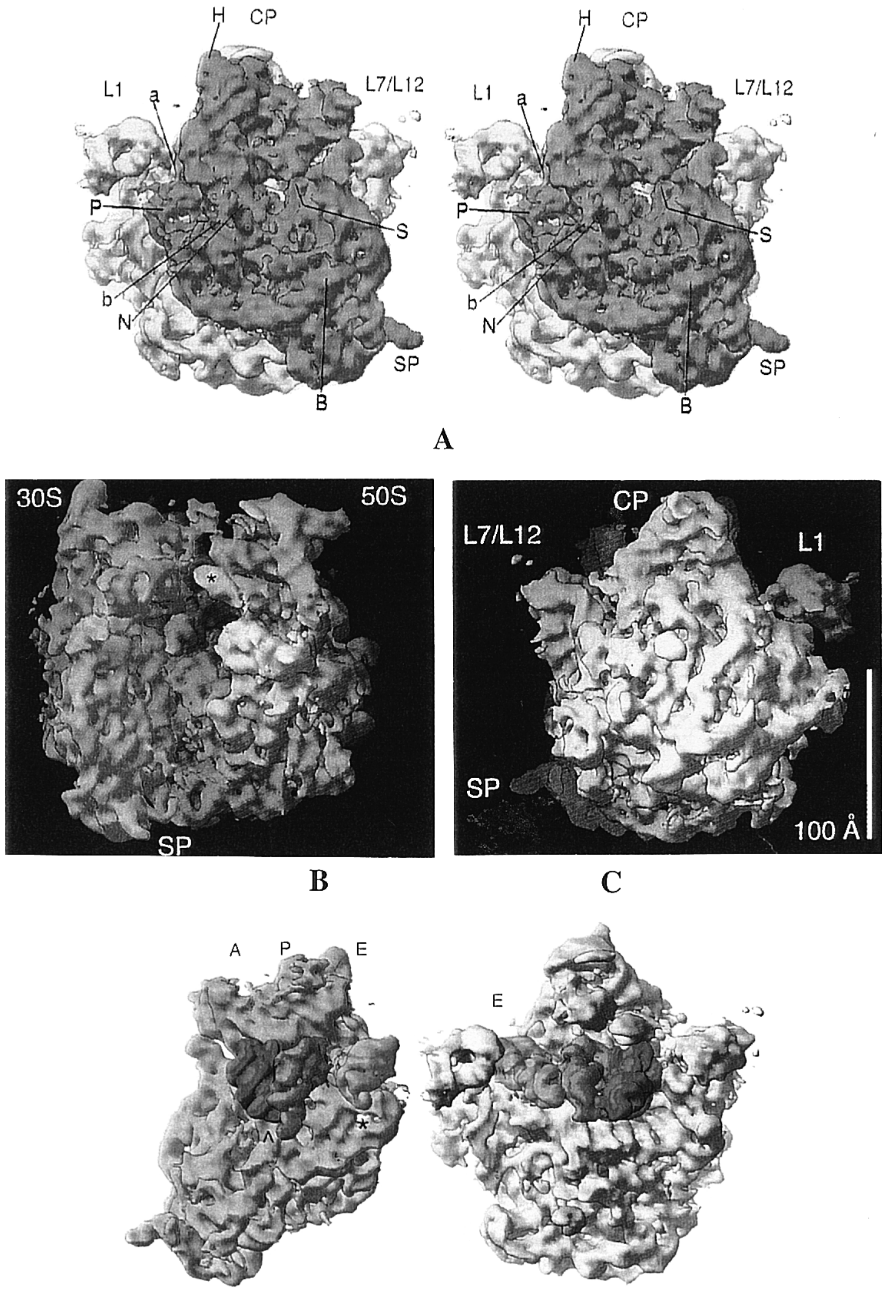

FIGURE 11 Electron density of the 70S ribosome. (A) Stereo view of the 0.78-nm electron density map from the solvent side of the 30S subunit (purple). The 30S subunit consists of the head (H) connected to the platform (P) and body (B). Additional features of the small subunit are the neck (N), spur (SP), shoulder (S), and contacts between the head and platform (a and b). The 50S subunit (gray) has the prominent features of the protein L1 stalk, central protuberance (CP), and L7/L12 region. (B) View of the ribosome rotated 90° about the vertical, with the L7/L12 region in the foreground. The A-site finger is marked with an asterisk. (C) View of the ribosome rotated 180° about the vertical axis compared with (A). (D) Model of three tRNAs bound to the ribosome, viewed toward the 30S interface (left) and toward the 50S interface (right). All three tRNA surfaces are electron density maps calculated to 0.78-nm resolution. The 3′-CCA end of the A-site tRNA (marked with ∧) is not modeled for lack of electron density to constrain its position. The 3′-CCA end of the E-site tRNA (asterisk) is buried at the base of the L1 stalk and cannot be distinguished at the present resolution. A, P, and E denote A, P, and E site, respectively. The peptidyl transferase center is located on the large subunit near the lower contact point of the A- and P-site-bound tRNAs. [From Cate, J. *et al.* (1999). *Science* **285,** 2095–2104. With permission.]

be favored, which may act either to stabilize or destabilize the mRNA, according to the needs of the cell, by determining its susceptibility to degradative enzymes (often termed *trans*-acting factors).

Whereas the structure of mRNA determines its susceptibility to degradative enzymes, the detailed mechanisms are complex. In prokaryotes, the enzymes involved include two endonucleases (RNase E and RNase III) and two exonucleases (polynucleotide phosphorylase and RNase II). Other nucleases may be active in particular cases such as phage infection. In eukaryotes, a major pathway involves removal of the 3′ poly(A) tail (deadenylation), followed by removal of the 5′ cap, which renders the mRNA susceptible to rapid endonucleolytic degradation in the $5' \rightarrow 3'$ direction.

B. Control by Interaction of Proteins with mRNA

Throughout the ribosome cycle, dynamic protein–mRNA interactions are functionally important in the initiation, elongation, and termination of polypeptide synthesis. In addition, more stable associations between proteins and mRNAs have been observed, particularly in eukaryotic cells. These messenger ribonucleoprotein complexes (mRNPs) occur both in polyribosomes and free in the cytosol, some of the latter being either temporarily or permanently unavailable for translation. Thus, protein-mRNA interactions contribute to the efficiency with which mRNAs are translated.

Some proteins, such as the poly(A)-binding protein (p78), are present in most if not all mRNPs, whereas others appear to be cell specific and mRNA selective. In unfertilized sea urchin eggs and *Xenopus* oocytes, for example, untranslated messenger is sequestered by association with proteins that prevent translation until later stages of development. Duck reticulocytes contain globin mRNP, which cannot be translated *in vitro*, whereas the mRNA obtained by deproteinizing the complex can be translated, showing that in this case translation is prevented by the mRNP proteins.

Formation of a site-specific mRNA-protein complex is involved in the translational control of the biosynthesis of ferritin, an iron storage protein, which is stimulated in response to the presence of iron. In this instance, a cytoplasmic repressor protein of 85 kDa binds to a highly conserved 28-nucleotide stem–loop structure in the 5′ untranslated region of ferritin mRNAs in the absence of iron. In the presence of iron, the protein dissociates from the mRNA, which is then available for translation. A similar loop motif occurs in the 3′ untranslated region of transferrin receptor mRNA, which is also subject to translational control by an iron-responsive repressor.

During the cell cycle, histone mRNA is destabilized after completion of DNA replication, resulting in a 30- to 50-fold decrease. This change appears to be due to an increase in the level of free histones, which form a complex with histone mRNA by interaction with a stem–loop structure at the extreme 3′ terminus. Formation of this histone–histone mRNA complex is thought to activate a ribsome-associated $3' \rightarrow 5'$-exonuclease, which degrades the histone mRNA. During the S phase, newly synthesized DNA binds free histones to form nucleosomes, thus preventing the degradation of histone mRNA at this stage of the cell cycle.

Specific regulation of gene expression at the level of translation also exists in prokaryotes. For example, the synthesis of *E. coli* threonyl-tRNA synthetase is negatively autoregulated by an interaction of the tRNA-like leader sequence of its mRNA with the synthetase, which inhibits translation by preventing the binding of ribosomes. The synthetase is displaced from the mRNA by $tRNA^{Thr}$, which thus acts as a translational antirepressor. This regulatory mechanism allows the cell to maintain a balance between the tRNA synthetase and its cognate tRNA.

Similarly, there is a mechanism used to control the synthesis of proteins encoded by a polycistronic mRNA. In this case, selective binding of the ribosomal protein to the region of the mRNA involved in the initiation of translation leads to the regulatory protein controlling both its own synthesis and that of other ribosomal proteins. A specific example is the role of ribosomal protein S4, which acts as a translational repressor of four ribosomal proteins (S4, S11, S13, and L17). Protein S4 appears to function as a repressor through an unusual "pseudoknot" linking a hairpin loop upstream of the ribosome-binding site with sequences 2 to 10 codons downstream of the initiation codon. (A pseudoknot structure contains intramolecular base pairs between base residues in the loop of a stem–loop structure and distal complementary regions of the RNA.) Stabilization of this structure by S4 would prevent the binding of ribosomes, and this control mechanism may contribute to the coordinated synthesis of the different ribosomal proteins required for ribosome assembly.

C. Control by mRNA Structure

The secondary structure of some eukaryotic mRNAs regulates translation by a mechanism involving a ribosomal frameshift which gives rise to a directed change of the translational reading frame to allow the synthesis of a single protein from two or more overlapping genes by suppression of an intervening termination codon. Several retroviruses use this mechanism to move from one reading frame to another in the expression of the viral

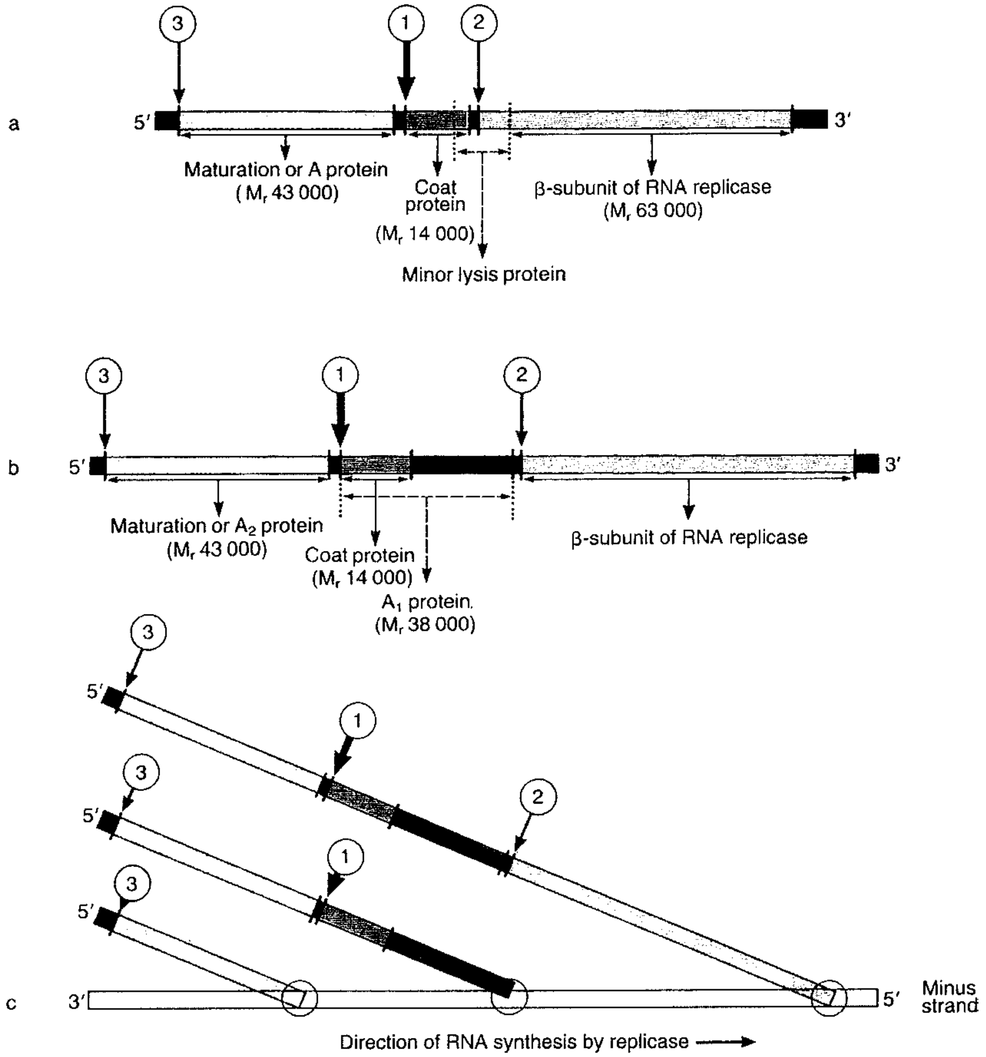

FIGURE 12 Translational control of bacteriophage synthesis. (a) Arrangements of MS2 and f2 bacteriophage cistrons. (b) Arrangement of Q_β bacteriophage cistrons. (c) Replicative intermediate synthesizing new plus strands on the complementary minus strand copy of the original bacteriophage RNA. The ribosome binding sites (RBS) are indicated by the numbered arrows. The major RBS (1) binds ribosomes efficiently but can be blocked by ribosomal protein S1. The secondary RBS (2) becomes available only after translation of at least part of the coat protein cistron by ribosomes. RBS 3 is masked by the secondary structure of native bacteriophage RNA but is accessible in nascent RNA (c) or *in vitro* when the secondary structure is destroyed. Noncoding regions are shown in black upstream from the RBS sites and at the 3-end. [From Arnstein, H. R. V., and Cox, R. A. (1992). "Protein Synthesis," Oxford University Press, Oxford. With permission.]

RNA-dependent DNA polymerase. Other example of the operation of such a frameshift include the synthesis of the reverse transcriptase enzymes of several retrotransposons, such as the yeast Ty1. The mechanism of changing the reading frame involves "slippery" sequences and a complex folding of the mRNA into a structure termed a pseudoknot.

In *E. coli*, translational control of the three cistrons of Q_β, f2, and related bacteriophage RNAs (Fig. 12) accounts for the synthesis of coat protein:replicase:A protein in the approximate ratio of 20:5:1. These quantitative differences are due to the differential and independent initiation of translation at each cistron as a result of differences in the secondary structure of the mRNA initiation sites. Furthermore, *in vivo* there is a delay in the synthesis of coat protein and this temporal control involves translational repression of the cistron by ribosomal protein S1. In addition, S1 functions as one of the subunits of the f2 RNA replicase; therefore, association of the newly synthesized translation product of the f2 replicase cistron with S1 will favor its dissociation from the phage RNA, thus allowing translation of the coat protein cistron to start.

D. Control by Modification of Translation Factor Activity

1. Initiation Factors

In eukaryotes, initiation of protein synthesis is inhibited by phosphorylation of the initiation factor eIF-2. Phosphorylation is stimulated by the lack of haem or the presence of double-stranded RNA. Two different protein kinases capable of phosphorylating the alpha subunit of eIF-2 have been characterized. One enzyme, called the haem-controlled repressor (HCR) or haem-regulated inhibitor (HRI), is a cytoplasmic protein (95,000 Da) that becomes activated by phosphorylation. The other kinase (67,000 Da) is activated by phosphorylation in the presence of double-stranded RNA. Thus, a cascade of protein phosphorylation is involved. Phosphorylated eIF-2 is unable to exchange GDP for GTP, which prevents it from functioning in the binding of initiator tRNA to ribosomes.

Conditions other than lack of haem (e.g., heat shock, serum deprivation, or the presence of oxidized glutathione), which are known to inhibit protein synthesis, also give rise to the phosphorylation of eIF-2α. Conversely, the activity of eIF-4F is decreased by dephosphorylation of the 24-Da subunit (see Table III) rather than by phosphorylation. Thus, a number of different kinases and phosphatases are involved in modulating the activities of different factors.

Small RNAs may also be involved in regulating the translation of mRNA in eukaryotic cells. Of the stimulatory RNAs, the best characterized is a small RNA of about 160 nucleotides, which accumulates in cells after infection with adenovirus. This virus-associated RNA, VA-RNA_1, which is required to maintain general protein synthesis, acts by inhibiting the phosphorylation of the alpha subunit of initiation factor eIF-2.

2. Other Translation Factors

A Ca^{2+}/calmodulin-dependent protein kinase phosphorylates eEF-2. The phosphorylated factor appears to be inactive and moreover also inhibits the activity of the nonphosphorylated factor. Dephosphorylation of the factor by phosphatase restores its activity.

E. Effects of Antisense Polynucleotides

Antisense RNAs, which are polynucleotides with base sequences complementary to messenger RNAs, have been found in both prokaryotes and eukaryotes. Natural antisense RNAs are not common but synthetic RNAs directed at specific targets have been widely studied. It has been demonstrated that they can function as inhibitors of messenger RNA translation. In prokaryotes, the most effective inhibitors appear to have a base sequence complementary to the 5′ leader region, including the Shine–Dalgarno sequence, which is involved in the binding of mRNA to the small ribosomal subunit. In eukaryotes, translation of mRNA is inhibited by polyribonucleotides complementary to the 5′ untranslated region of mRNA, indicating a direct effect on initiation as in prokaryotes. Polynucleotides complementary to the 3′ untranslated region of mRNA also inhibit translation in some cells, and this effect may be due to destabilization of mRNA by ribonucleases specific for double-stranded RNA. The effect of antisense polynucleotides is not restricted to translation of mRNAs, but transcription as well as the processing of transcripts may also be inhibited.

F. Availability of Amino Acids, tRNA Abundance, and Codon Usage

1. Amino Acids

Polypeptide synthesis depends on an adequate supply of tRNAs charged with the 20 protein amino acids and appropriate interactions between their anticodons and the codons of mRNA. Peptide chain elongation is decreased or inhibited by lack of amino acids or other conditions giving rise to an imbalance or deficiency in aminoacyl-tRNAs.

2. Abundance of tRNAs and Codon Usage

Different tRNAs are present in the cytosol in unequal amounts, and elongation rates are slower at codons corresponding to rare tRNA species.

The existence of synonymous codons raises the question of preferential use of some codons and its possible significance in relation to translational efficiency and control. In some bacteria (e.g., *Pseudomonas aeruginosa*, which has a high content of G + C, 67.2%, in DNA), the most common codons are those with the strongest predicted codon–anticodon interaction—that is, G + C base pairs—but this preference is not universal and, for example, does not apply to *E. coli*, which has a lower proportion of G + C (50%). Although codon usage may play a part in determining elongation rates, it is probably of less importance in translational control than the secondary structure of mRNA in relation to the rate of initiation of protein synthesis.

G. Modulation of Ribosome Activity

Specific ribosomal components have an important function in relation to the fidelity of protein synthesis. Thus, in *E. coli* ribosomal protein S12 determines the accuracy of codon–anticodon interactions and modulates the

translational error frequency in the presence of the antibiotic streptomycin.

To what extent reversible modifications of ribosomal constituents are involved in translational control of protein synthesis is uncertain. Although phosphorylation of ribosomal protein S6 increases with cell proliferation, it is not known whether this change is directly related to the accompanying increase in protein synthesis by an effect on the translation rate.

H. Ribosome-Inactivating Proteins

Many molds and plants produce toxins, which are protective reagents, termed ribosome-inactivating proteins (RIPs), directed at particular cells and their ribosomes. These toxins are classified as either type I or type II RIPs according to the number of polypeptide chains.

Type I RIPs comprise a single polypeptide chain; for example, α-sarcin, an extracellular cytotoxin produced by *Aspergillus giganteus*, consists of a single chain of 150 amino acid residues.

Type II RIPs comprise two polypeptide chains, A and B. The A chain has the ability to inactivate ribosomes, and the B chain is a galactose-specific lectin responsible for the entry of the toxin into the target cell. Ricin, which is isolated from castor beans, is representative of type II RIPs.

Ribosomes are inactivated as a result of the RNA N-glycosidase activity of RIPs. These toxins have different specificities for particular cells and ribosomes. However, the target site for all RIPs is an adenylate residue (position 2660 in the *E. coli* 23S rRNA sequence) located within a highly conserved sequence of 12 nucleotides ($5'_{2654}$AGUACGAGAGGA$_{2665}3'$). Cleavage of the GpA$_{2660}$ internucleotide bond or depurination of A$_{2660}$ is sufficient to inactivate the ribosome. The target residue is located in the loop region of a stem–loop element of secondary structure which is termed the α-sarcin stem–loop. Thus, the target adenylate is either directly or indirectly essential for ribosome function. The α-sarcin stem–loop is known to be important for binding elongation factors EF-Tu and EF-G to the ribosome. RIPs have attracted interest as active components of reagents directed at particular targets such as cancer cells.

VI. CONCLUDING REMARKS

The control of protein synthesis, either by regulation of the amount of mRNA available for translation or by the efficiency with which it is translated, is important in cell growth and development as a factor determining the level of cellular and extracellular proteins. Subversion of this control occurs in cells infected by viruses when the viral nucleic acid uses the protein-synthesizing machinery of the host cell and thereby changes normal cell metabolism in favor of the synthesis of viral proteins needed for the production of virus progeny.

The polypeptide chains of all proteins are synthesized by the process described above. This mechanism gives rise to primary polypeptide chains, which are often further modified—for example, by cleavage into smaller peptides, by structural modification of selected amino acid residues, by splicing of the polypeptide chain, or by the formation of covalent bonds between polypeptide chains. Some of these secondary modifications are related to the correct folding of polypeptide chains and to the production of active enzymes or peptide hormones from inactive precursors (e.g., insulin from proinsulin). Also, the transport of proteins within the cell or the secretion of extracellular proteins is often linked to structural changes in polypeptide chains either during or after completion of synthesis.

SEE ALSO THE FOLLOWING ARTICLES

Biomaterials, Synthesis, Fabrication, and Applications • Cell Death (Apoptosis) • Chromatin Structure and Modification • DNA Testing in Forensic Science • Gene Expression, Regulation of • Nucleic Acid Synthesis • Protein Folding • Protein Synthesis • Ribozymes

BIBLIOGRAPHY

Al-Karadaghi, S., Kristensen, O., and Liljas, A. (2000). "A decade of progress in understanding the structural basis of protein synthesis," *Progr. Biophys. Mol. Biol.* **73,** 167–193.

Arnstein, H. R. V., and Cox, R. A. (1992). "Protein Biosynthesis," Oxford University Press, London.

Ban, N., Nissen, P., Hansen, J., Moore, P. B., and Steitz, T. A. (2000). "The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution," *Science* **289,** 905–920.

Carter, A. P., Clemons, W. M., Brodersen, D. E., Morgan-Warren, R. J., Wimberly, B. T., and Ramakrishnan, V. (2000). "Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics," *Nature* **407,** 340–349.

Garrett, R. A. *et al.*, eds. (2000). "The Ribosome: Structure, Function, Antibiotics and Cellular Interactions," American Society for Microbiology, Washington, D.C.

Green, R., and Noller, H. F. (1997). "Ribosomes and translation," *Annu. Rev. Biochem.* **66,** 674–716.

Grunberg-Manago, M. (1999). Messenger RNA stability and its role in control of gene expression in bacteria and phages," *Annu. Rev. Genetics* **33,** 193–227.

Hardesty, B., and Kramer, G. (2001) "Folding of a nascent peptide on the ribosome," *Progr. Nucleic Acid Res. Molec. Biol.* **66,** 41–66.

Herr, A. J., Atkins, J. F., and Gesteland, R. F. (2000). "Coupling of open reading frames by translational bypassing," *Annu. Rev. Biochem.* **69,** 343–372.

Hershey, J. W. B., Mathews, M. B., and Sonenberg, N., eds. (1996). "The Pathway and Mechanism of Eukaryotic Protein Synthesis," Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Hill, W. E., Dahlberg, A., Garrett, R. A., Moore, P. B., Schlessinger, D., and Warner, J. R. (1990). "The Ribosome, Structure, Function and Evolution," American Society for Microbiology, Washington, D.C.

Ibba, M., and Söll, D. (2000). "Aminoacyl-tRNA synthesis," *Annu. Rev. Biochem*. **69,** 617–650.

Matheson, A. T., Davies, J., Hill, W., and Dennis, P. (1996). "Frontiers in Translation," National Research Council of Canada, Ottawa.

Nierhaus, K. H., Subramanian, A. R., Erdmann, V. A., Franceschi, F., and Wittmann-Liebold, B., eds. (1994). "Translational Apparatus: Structure, Function, Regulation and Evolution," Plenum Press, New York.

Nissen, P., Hansen, J., Ban, N., Moore, P. B., and Steitz, T. A. (2000). "The structural basis of ribosome activity in peptide bond synthesis," *Science* **289,** 920–930.

Rodnina, M. V., Savelsbergh, A., and Wintermeyer, W. (1999). "Dynamics of translation on the ribosome: molecular mechanics of translocation," *FEMS Microbiol. Rev*. **23,** 317–333.

Shine, J., and Dalgarno, L. (1974). The 3′-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proceedings of the National Academy of Sciences of the United States of America* **71,** 1342–1346.

Wimberly, B. T., Brodersen, D. E., Clemons, Jr., W. M., Morgan-Warren, R. J., Carter, A. P., Vonrhein, C., Hartsch, T., and Ramakrishnan, V. (2000). "Structure of the 30S ribosomal subunit," *Nature* **407,** 327–339.